

IMPLEMENTATION OF A GIS WITH ANALYTICS FOR POLLUTION DATA

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Aakanksha Rastogi

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Software Engineering

March 2015

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Implementation of a GIS with analytics for pollution data

By

Aakanksha Rastogi

The Supervisory Committee certifies that this *disquisition* complies with
North Dakota State University's regulations and meets the accepted standards
for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Kendall Nygard

Chair

Dr. Gursimran Walia

Dr. Limin Zhang

Approved:

03/16/2015

Date

Brian Slator

Department Chair

ABSTRACT

Particulate matter called pm2.5 refers to harmful microscopic liquid or solid matter of diameter 2.5 microns suspended in the air. To effectively analyze and visualize data for pollution category pm2.5, a Geographic Information System (GIS) was created. This application, named the Pollution Analysis System (PAS), was designed and implemented to help the user visualize the data in terms of graphs, charts and maps. The interactive visualizations are aimed to enhance data readability and give meaning to the raw data. Correlation techniques were implemented, including calculations of the Pearson correlation coefficient, the use of the t-distribution to test the validity of correlation between the two data sets. Also, the p value was calculated to provide the confidence level of the correlation hypothesis and its significance.

ACKNOWLEDGEMENTS

I would like to express my immense gratitude towards Dr. Kendall Nygard for extending the opportunity to pursue my research work under him. I would also like to thank Dr. Gursimran Walia and Dr. Limin Zhang for their valuable presence as my committee members.

I would also like to thank my family for their blessings and providing me with constant support and motivation and having faith in me. Also, my friends and Karan without whose support this research would not have been made possible.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES.....	ix
LIST OF APPENDIX TABLES.....	x
CHAPTER 1. INTRODUCTION.....	1
1.1. Purpose.....	1
1.2. Motivation.....	2
1.3. Problem Statement.....	2
1.4. Paper Organization.....	3
CHAPTER 2. LITERATURE REVIEW.....	6
2.1. Background.....	6
2.2. Existing Work.....	7
CHAPTER 3. DESIGN AND ARCHITECTURE.....	10
3.1. Overview.....	10
3.2. Subsystem Decomposition.....	10
3.3. External Interface Requirements.....	15
3.3.1. User Interfaces.....	15
3.3.2. Hardware Interfaces.....	16
3.3.3. Software Interfaces.....	17
3.3.4. Communications Interfaces.....	17
3.4. Persistent Data Management.....	17
CHAPTER 4. IMPLEMENTATION.....	22
4.1. Data Analysis And Visualization.....	22

4.2. Pollution Analysis System	27
CHAPTER 5. PERFORMANCE.....	43
CHAPTER 6. TESTING.....	46
CHAPTER 7. DISCUSSION.....	86
7.1. Strengths	86
7.2. Limitations	88
CHAPTER 8. CONCLUSION AND FUTURE WORK	90
8.1. Conclusion	90
8.2. Future Work	91
CHAPTER 9. REFERENCES	92
APPENDIX.....	94

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Use Cases	11
2. User Interface pages.....	16
3. Test for year values on User Home page	46
4. Test for year value 2011 on Visual Representation of Emissions Data display	48
5. Test for year value 2011 on Visual Representation of Emissions Data display	49
6. Test for category tier dropdown on Visual Representation of Emissions Data display	52
7. Test for sum of emissions for all states for a category tier for year 2011.....	55
8. Test for sum of emissions for all states for a category tier for year 2008.....	56
9. Test for selected year value 2011 on More Visualizations for Emissions data	58
10. Test for select year value 2008 on More Visualizations for Emissions data	60
11. Test for State dropdown.....	62
12. Test for Category Tier dropdown on More Emissions of Emissions Data display	64
13. Test for county with maximum emission in year 2011.....	67
14. Test for county with maximum emission in year 2008.....	68
15. Test correlation coefficient for same states and same year.....	70
16. Test correlation coefficient for same states and different years	71
17. Test for negative correlation for same states and different years	73
18. Test for positive correlation for different states and different years.....	74
19. Test for negative correlation for different states and different years.....	76
20. Test for positive correlation for different states and same year	77
21. Test for negative correlation for different states and same year.....	79
22. Test for positive correlation for different states and same year.....	80
23. Test for number of days with pm2.5 and number of days with AQI for a state	82

24. Test for average AQI data for pm2.5 for highlighted state..... 84

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Use case diagram - Admin user	12
2. Use case diagram - Regular user.....	13
3. Use case diagram - Emissions data.....	14
4. Use case diagram - Air Quality Index data.....	15
5. Admin User.....	28
6. User home	29
7. Visual Representation of Emissions Data.....	30
8. More Visualizations for Emissions data	32
9. Correlation of Emissions Data	33
10. Probability (P) Value T test Calculator.....	34
11. Motion Chart (View 1).....	35
12. Motion Chart (View 2).....	37
13. Motion Chart (View 3).....	38
14. Table (View 1)	39
15. Table (View 2)	41
16. Geo Map.....	42

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
A1. Use case to describe the Admin user Home page	94
A2. Use case to describe the upload functionality for the Air Quality Index data on the Admin user Home page	95
A3. Use case to describe the upload functionality for the Emissions data on the Admin user Home page.....	96
A4. Use case to describe the User Home page	97
A5. Use case to describe the Visual Representation of Emissions data page.....	98
A6. Use case to describe More Visualizations for Emissions data page	99
A7. Use case to describe the Correlation of Emissions data page	100
A8. Use case to describe the Motion Chart for Air Quality Index data page	101
A9. Use case to describe the Table for Air Quality Index data page.....	102
A10. Use case to describe the Geo Map for Air Quality Index data page	103

CHAPTER 1. INTRODUCTION

Pollution arises from various sources and pollutants constituting it has varied impacts depending upon its size and composition. Understanding pollution caused by particulate matter holds utmost importance if we are living in a highly polluted place. This chapter describes the purpose of building an application which can effectively picturize the impact of pollutants consisting of particulate matter. It will also emphasize on the main focus of this research work. Chapter One discusses the purpose of building PAS application and why is it useful.

1.1. Purpose

Particulate matter is categorized in accordance to their size since they pose different health effects depending on their different diameters. It is a general term which is used for a combination of liquid droplets and solid particles suspended in the air. This may include, and are not limited to, smokes, aerosols, fumes, dust, pollen and ash. pm2.5 refers to particulate matter whose diameter is 2.5 microns or less. It is also referred to as fine particulate matter or respirable particles since their diameter is very small and it can easily penetrate the respiratory system. pm2.5 has been responsible for affecting health to the maximum extent which may occur over both short periods (such as a day) and longer periods (a year or more). If exposed, it can result in hospital admissions, several serious health issues, and even premature death. People which are the most sensitive to pm2.5 include the ones with cardiovascular or lung disease, asthma, children and elderly people [9].

The primary reasons behind the formation of pm2.5 material are fuel combustion (for example, power generation, motor vehicles, industrial facilities, residential fire places, wood stoves and agricultural building) and chemical reactions in the atmosphere. pm2.5 is also

accountable for environmental effects such as damage to vegetation, soiling, reduced visibility and corrosion [9].

Pollution data in category pm2.5 is so hazardous that understanding its patterns and trend has become a necessity in today's world. The purpose is to gather the data of category pollutant pm2.5, process the data, analyze and visualize the data in terms of graphs, charts and maps. With growing urbanization and industrialization, pollution and a variety of pollutants causing it, especially, the ones resulting in category pm2.5, have gone way out of hand. While some places share the joy of clean air and a healthy environment, there are places which are severely struck with high levels of pollution including pollutants dispensing higher amounts of category pm2.5.

1.2. Motivation

While manufacturing goods to facilitate basic needs, the waste generated in the process and the pollutants released through could not really catch much of an attention and hence was not taken care off. Literacy among people on pollutants category pm2.5 is very less. Attempting to generate a better understanding of pollutants in category pm2.5 and to create an awareness to take precautionary measures motivated me to design Pollution Analysis System (PAS). In order to understand the damage done by the pollutants and raising awareness about the US states or cities which have very high levels of pollution, it was necessary to come up with an interactive system which can provide user with a walk through of all these pollution causing culprits.

1.3. Problem Statement

The prime focus of this research concerns the application of Geographic Information System (GIS) augmented by the use of data analytics with charts and graphs other than just maps. This includes designing a GUI based system or PAS which would interface analytical and

statistical information using pollution data in category pm2.5. Correlation technique used in PAS will help the user compare the impact of pollution caused by pollutant in category pm2.5 among different states for same year or same states for different years. Correlation can be calculated depending upon the state and the year choices user makes. The system will also display the sum of all the emissions from all the sources of pollution (or category tier types) from the two states for same or different years. With the help of visual representation in the bar chart, the user will be able to tell if the pollution levels increased or decreased going from one year to another. If they increase, the user can deduce if what could have been done better to reduce the levels of pollution. Similarly, the user will also be able to compare the pollution data amongst several other geographic locations or different states for different years. In the event that other states' pollution levels are lower than the state user is interested in; it will greatly help the user understand that other states must have been doing something different to have their pollution levels under control. The user can also deduce or ponder upon the healthy lifestyle approaches other states may have been following. This GIS will allow us to analyze, interpret and understand the pollution data in category pm2.5. It will also help us with visualizing the data in terms of graphs, charts and maps with the use of data analytics.

1.4. Paper Organization

This paper is organized as and composed of seven chapters which are described as follows. Chapter One discusses the purpose of building PAS application and why is it useful. This chapter also includes the problem statement which discusses the main focus of this research work.

Chapter Two discusses the background of this PAS application and existing work which has been done in this area. It also elaborates on the comparison of PAS application to the related work which has been done in this direction.

Chapter Three discusses about the design and architecture of this application, how the application is designed, engineered and the tools and language choices which were opted for. This chapter also describes why those tools and language choices were made and briefs on its comparison to other tools and languages. Also included in this chapter is the subsystem decomposition which briefs on the use cases and the basic flow of the application. The persistent data management part of this chapter elaborates upon the database which was selected during the development of this application and the data structures involved.

Chapter Four is the main Implementation section and discusses about how the system application was implemented and what is it based upon. It elaborates on the data sources and how the data was analyzed and visualized and made it easier for the user to comprehend. This also discusses the correlation techniques which were used to further analyze the data and correlate to have a better understanding of the impact of pollution in category pm2.5. Other section of this chapter elaborates on the basic prototype of the PAS application and describes all the charts, maps, graphs and tables used to visualize the data.

Chapter Five discusses the performance of the application.

Chapter Six discusses the testing approaches which were chosen to test this application to its entirety. This also includes the testing measures and standards used to test each and every module of the PAS application.

Chapter Seven discusses the strength and limitations of this application.

Chapter Eight discusses the conclusion of this application and research work and the future work that can be done in this direction to further enhance this application.

Chapter Nine enlists all the references which were taken while building this application and writing this paper.

An Appendix focuses on the working of the Pollution Analysis System (PAS) in the form of detailed use cases.

CHAPTER 2. LITERATURE REVIEW

Before creating Pollution Analysis System (PAS) application, it is important to gather information from the previous work which was done to understand pm2.5, its impact on health and its contribution towards the growing pollution. The focus of this chapter is to discuss the background of PAS application and existing work which has been done in this area and to compare this application to any related work which has been done before.

2.1. Background

Pollution in category pm2.5 (particulate matter 2.5) refers to the pollutants or the particles less than 2.5 μm in diameter. Studies in the United States and everywhere else on pollutants from this category have suggested that the particle size this less is responsible for morbidity and mortality [11]. Pollutants of particle size pm2.5 get lodged into the lungs very easily and have been known to cause a variety of lung diseases and other respiratory problems. pm2.5 has adverse health effects as compared to pm10 and has stronger associations with the total suspended particles [13]. Besides the adverse health effects of pm2.5 pollutants, it has the ability to stay suspended in the atmosphere for longer durations of time and travel from one place to another. For example, pm2.5 pollutants found in the atmosphere over Fargo can travel up to the atmosphere over Grand Forks. In their study on Chemical Characteristics of PM2.5 and PM10 in Haze–Fog Episodes in Beijing, Sun et al compared the characteristics and effects of pm2.5 and pm10 pollutants and reported that “the sources of PM2.5 were more from areas on the regional scale due to its tendency for long-range transport, while PM10 was more limited to the local sources” [17].

In their research work, Engel-Cox et al. deduced a visualization technique which will put together back trajectory analysis, ground-based data and satellite data together [8]. They believe

these sensors are capable of monitoring air quality data in rural or remote regions with minimal to no ground-based monitoring network. My research will use a similar approach as part of future work except that the sensors won't be satellite sensors.

Raffuse et al. developed a spatial mapping tool to characterize relationships between downward air quality observation and air pollutant emissions [15]. This helped them describe wind trajectory probability distributions and spatial distribution of emissions.

It can be a challenge to visualize the data which is collected over a period of time and at multiple locations and resulting in multivariate time series which are geographically indexed. In his paper, Peng has made an attempt to visualize multivariate time series data by presenting `mvtsplot` function [14].

Snyder et al. focused upon the importance of monitoring air pollution and that how different types of sensors and their usage impacts the quality of monitored data. They have tried to approach sensors which are least expensive and monitor air pollution to a fine granularity so that it can aid people with health conditions sensitive to air pollution and several harmful pollutants that are a part of it. They have separated sensors into two main categories – one who measure the concentration of gas phase species and the others that measure either the particulate matter or various properties of particles [16].

2.2. Existing Work

Interactive visualization of pollution data in terms of graphs, charts and maps is important since it helps user compare the statistics and raw data visually. Air Quality and Pollution measurement team has been constantly monitoring the pollution data across several cities in the world and have reported it visually in terms of charts and maps. This includes displaying the pollution data for all the air pollutants such as pm2.5, pm10, O3, NO2, SO2 and CO. Their data

representations are frequently streamed to display the most recent data statistics and calculations to reveal the air quality index numbers for a particular city. Not only this, they have done a pretty fair analysis of the data in terms of color coding it and displaying the healthy, unhealthy, moderate or very unhealthy levels of current air quality. For example, by visualizing a live state of the air quality index of Beijing being ‘Very Unhealthy’, it can be concluded that it’s time for a mask [5]. Other visualizations that they provide include and are not limited to AQI, Air Pollution Level and its possible health implications.

A fair share of interactive visualizations of data in terms of motion charts and bubble charts has been done by Gapminder Foundation [10]. Gapminder is a ‘multivariate Javascript package’ [12]. It visualizes several animated statistics about different countries of the world. These animated statistics include – health among several cities, CO2 emissions in USA as compared to China, CO2 emissions since 1820, globalizations, human development trend from 1960 to 2001, last 40 years of development and urbanizations and much more. Gapminder is a Swedish based non-profit venture which was co-founded by Dr. Hans Rosling in 2005 [1]. In their paper, Cheng & Battista mentions Professor Rosling as a pioneer of motion charts [4]. Dr. Rosling has been practicing medicine and has collected and analyzed data related to health and disease. He has also researched social and global economic issues among countries which are in various development stages [4]. With the help of a team of graduate students, he invented and developed ‘Trendalyzer’ software which is used to create dynamic charts and graphs on statistical series of global health and economics [4].

Trendalyzer is visualization software which can convert statistical time series data into animated and interactive graphics [2]. The current version of Trendalyzer is a beta version and is a Flash application. This software visualizes the data in the form of interactive bubble charts with

the help of preloaded historical and statistical data [19]. In 2007, the Trendalyzer software was acquired by Google. Google took the Flash-based Motion Chart component of this software and made it available to public as open source software by 2008 [2]. Google made it public as their own modified version called Google Motion Chart which allows the user to work on with their own input data. The idea of visually representing the Air Quality Index data in terms of bubble chart or motion chart was inspired by Gapminder and Google Visualization API was used to create them. However, the type of data used, the way it is bound to the chart on the display and the data structures used in the skeleton of the application are different than the way Gapminder does it.

My research work and the PAS will be further enhanced by Dr. Nygard's prospective research ideas which propose fetching live data from pollution sensors. People participating in research or volunteering for research will carry sensors with them to record pollution data for criteria pollutant pm2.5. These volunteers or participants will be people going to school or work or running errands in their day to day life as well as passing through a high pollution zone. These people will be using either public modes of transportation or personal vehicles or some of them would either be walking. After they have the data captured through the sensors, they can record it or transfer it to their smart phones and through the smart phones this data can be uploaded to servers.

CHAPTER 3. DESIGN AND ARCHITECTURE

Design and architecture of the application is the next step after setting the focus of the research and gathering requirements and information. This chapter provides in depth information on how the application is designed and created, all the language choices which were made and the tools which were used to create the application. Efficient data management and the data structures involved are described in the persistent data management section of the chapter. The reason behind making the language and tools choices is also covered in this chapter. Apart from that, it also encompasses the basic flow of the application in terms of use cases.

3.1. Overview

The following proposed system aims to develop web application for facilitating interactive visualization of pollution data in category pm2.5. PAS will provide the end users with the in depth information on how to navigate through the system, view the pollution data for years 2008 and 2011 in terms of graphs, charts and maps, correlate the pollution data for a state or a city between years 2008 and 2011, visualize the air quality index for the years ranging from 1980 to 2013 in terms of motion charts and big table.

3.2. Subsystem Decomposition

The use cases associated with the PAS are as follows:

Table 1. Use Cases

N	Use case N	Brief Description
UC1	Admin user Home page	Admin user enters the Admin user Home page on the PAS
UC2	Admin user uploads Emissions data	Admin user uploads the Emissions data into the PAS
UC3	Admin user uploads Air Quality Index data	Admin user uploads the Air Quality Index data into the PAS
UC4	User Home page	Regular user enters the User Home page on the PAS
UC5	Visual Representation of Emissions data	User enters the Visual Representation of Emissions data page upon selecting a year from the year dropdown menu option on the User Home page to visualize the emissions data
UC6	More Visualizations for Emissions data	User enters the more Visualizations for Emissions data page upon clicking the hyperlink on the Visual Representation of Emissions data to visualize the emissions data further on a state level
UC7	Correlation of Emissions data	User enters the Correlation of Emissions data page upon clicking the hyperlink on the More Visualizations for Emissions data page to visualize the correlation between different states for same/different year or same state for different years
UC8	Motion Chart for Air Quality Index data	User enters the Motion Chart for Air Quality Index data page upon clicking the hyperlink on the User Home page to visualize the Air Quality Index data in the form of Motion Chart
UC9	Table for Air Quality Index data	User enters the Table for Air Quality Index data page upon clicking the hyperlink on the User Home page to visualize the Air Quality Index data in the tabular form
UC10	Geo Map for Air Quality Index data	User enters the Geo Map for Air Quality Index data page upon clicking the hyperlink on the User Home page to visualize the Air Quality Index data in the form of Geo Map



Figure 1. Use case diagram - Admin user



Figure 2. Use case diagram - Regular user



Figure 3. Use case diagram - Emissions data

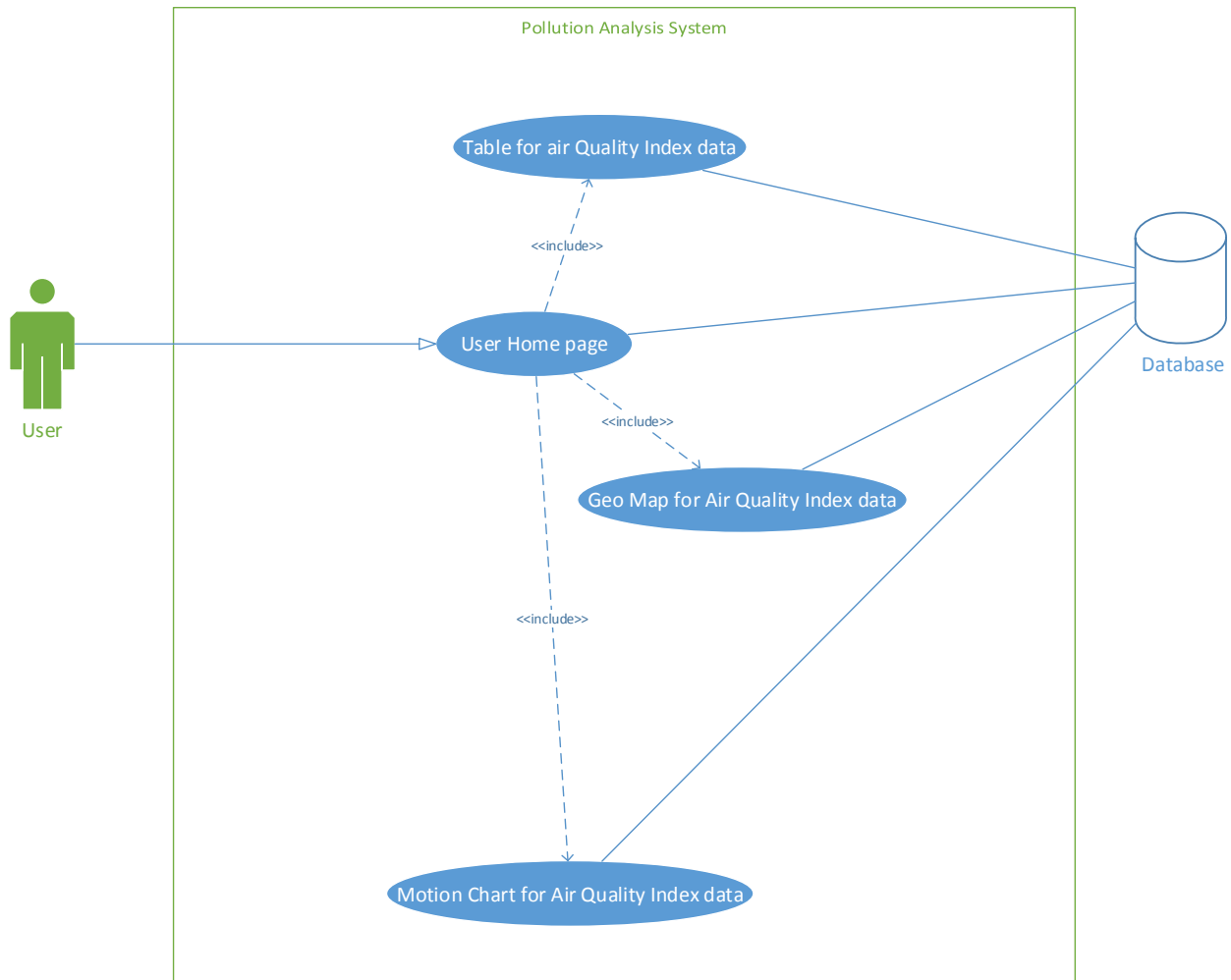


Figure 4. Use case diagram - Air Quality Index data

3.3. External Interface Requirements

3.3.1. User Interfaces

The User Interface pages are as follows:

Table 2. User Interface pages

Name of the page	Page Description
Home page	This page is an Admin home page which allows the Admin user of the application to be able to upload the data files to the system and the database. This data will be later viewed by the regular user of the PAS application.
UserHome page	This page allows the user to be able to select the year and visualize its data in the form of graphs, bar charts, motion chart, geo map and tabular form
Visual page	This page allows the user to be able to visualize the data in the form of bar chart. The user will have the option to be able to select a year and visualize its data for 14 category tiers or sources of pollution
Visualizations page	This page allows the user to be able to visualize the pollution data on a bar chart and view the geographic location of the city having highest amounts of pm2.5 levels on the map
Correlation page	This page allows the user to be able to correlate the level of pm2.5 between cities/same cities for years 2008 and 2011
MotionChart page	This page allows the user to visualize the trend of the levels of pm2.5 for the years ranging from 1980 to 2013. The user has the ability to either play the recording for the years worth of data or statically view the data
Geo Map page	This page allows the user to visualize the data in terms of geo chart and color codes different states on the basis of presence of levels of pm2.5. Upon hovering on each state, the interactive Geo Chart displays the name of the state and the average pm2.5 value.
Table for Air Quality Index data page	This page displays the air quality index data to the user in tabular form. The user will have the opportunity to make year selection and view the data accordingly.

Refer to the Appendix for more detailed description of these use cases.

3.3.2. Hardware Interfaces

The computers using the PAS should have the following hardware requirements:

- Microsoft Windows Professional, or Microsoft Windows XP, or Microsoft Windows Vista, or Microsoft Windows 7, or Mac, or Linux, or Unix-like OS
- 1.6 GHz Processor (32 bit or 64 bit) or higher
- RAM of 4GB or higher
- 1 GB or higher available hard disk space

3.3.3. Software Interfaces

The PAS is developed and implemented on ASP.NET. Being built on .NET framework, ASP.NET provides the application programming interface (API) and uses Microsoft's Internet Information Services (IIS) as a platform.

PAS uses a database which is configured using MySQL 5.6 and interfaces MySQL Workbench 6.0 CE. The data elements of the database are described in the Persistent Data Management section of this chapter.

3.3.4. Communications Interfaces

Since the system will be developed using ASP.NET, it will work with HTTP protocol and will use HTTP commands and policies to set a two-way communication and cooperation between browser and server. The user computers using this web-based application should have internet connection in them with accessibility to standard web browsers such as Mozilla Firefox, or Chrome, or Internet Explorer, or Safari.

3.4. Persistent Data Management

The database used while developing the system is MySQL 5.6. MySQL Workbench 6.0 CE uses open source tool MySQL DBMS as an interface for maintaining and managing the database. This also integrates well with Microsoft Visual Studio 2010.

The database consists of two tables – “datafromfile” and “airqualityindex”. Table “datafromfile” consists of all the emissions data pertaining to the states, respective counties and their corresponding category tiers for years 2008 and 2011. The columns from this table are –

- **uid** (this column contains a unique Id created for each row of data in the table),

- **file Id** (this column contains a unique file Id which is assigned to each new file added to the system by the Admin user)
- **fips** (this column contains a key or a code for each county belonging to a state; each county has its own fips code)
- **stips** (this column contains a code or an id for a state; each state has its own code or id)
- **ctyfips** (this column contains a code or an id for a county from a state; each city from a state has its own code or id)
- **state** (this column consists of state abbreviations)
- **countyName** (this column consists of all the counties belonging to their respective states)
- **tier** (this column consists of id's or codes for all of the 14 category tiers)
- **tierName** (this column consists of the tier names for their corresponding tier id's or codes)
- **pollutantCode** (this column contains the pollutant code; the pollutant code for this file being pm2.5)
- **pollutantName** (this column contains the pollutant name corresponding to the pollutant code; the pollutant name for this file being pm2.5 primary)
- **inventoryYear** (this column contains the inventory year for the emission data)
- **emissions** (this column contains the emission data in tons)
- **dataTimeStamp** (this column stores the date and time stamp of the time when the file was uploaded to the system by the Admin user)

Table “airqualityindex” consists of air quality data pertaining to the states, respective counties and their corresponding category tiers for years 1980 and 2013. This air quality data

also includes information about when the air quality was good, moderate, unhealthy for sensitive groups, unhealthy and very unhealthy. The columns from this table are –

- **uid** (this column contains a unique Id created for each row of data in the table),
- **file Id** (this column contains a unique file Id which is assigned to each new file added to the system by the Admin user)
- **cbsaCode** (this column contains the state abbreviations)
- **cbsaCity** (this column contains the names of the city)
- **numberOfDaysWithAQI** (this column represents the number of days in the year having an Air Quality Index. This is the number of days on which measurements from any monitoring site in the city were reported [3].)
- **good** (this column represents the number of days in the year having AQI value 0 through 50 [3])
- **moderate** (this column represents the number of days in the year having AQI value 51 through 100 [3])
- **unhealthyForSensitiveGroups** (this column represents the number of days in the year having AQI value 101 through 150 [3])
- **unhealthy** (this column represents the number of days in the year having AQI value 151 through 200 [3])
- **veryUnhealthy** (this column represents the number of days in the year having AQI value 201 or higher. This includes the AQI categories very unhealthy and hazardous [3])
- **AQIMaximum** (this column represents the highest daily AQI value in the year [3])
- **AQI90thPercentile** (this column represents the 90 percent of daily AQI values during the year were less than or equal to the 90th percentile value [3])

- **AQIMedian** (this column represents half of daily AQI values during the year were less than or equal to the median value, and half equaled or exceeded it [3])
- **numberOfDaysCO** (this column represents the number of days pollutant ‘CO’ measured was the main pollutant)
- **numberOfDaysNO2** (this column represents the number of days pollutant ‘NO2’ measured was the main pollutant)
- **numberOfDaysO3** (this column represents the number of days pollutant ‘O3’ measured was the main pollutant)
- **numberOfDaysSO2** (this column represents the number of days pollutant ‘SO2’ measured was the main pollutant)
- **numberOfDaysPM25** (this column represents the number of days pollutant ‘pm2.5’ measured was the main pollutant)
- **numberOfDaysPM10** (this column represents the number of days pollutant ‘pm10’ measured was the main pollutant)
- **year** (this column represents the year for which all of the air quality index data was reported to the Air Quality Index data)
- **dateTimeStamp** (this column stores the date and time stamp of the time when the file was uploaded to the system by the Admin user)

For each of the air pollutant measured, a daily index value was calculated. Out of all of these index values, the highest of those values is the AQI value. The pollutant which is responsible for the highest index value is the “Main Pollutant” [3].

Whenever the user or the Admin user downloads any file, it will be saved on the server. The server would be where the code is running or where the application is being deployed. Also,

from wherever the file is being uploaded to the system, it will be saved to the server. This will help restore the data in the events of possible data loss due to database crash or unavailability of backup data.

CHAPTER 4. IMPLEMENTATION

Pollution Analysis System (PAS) application is a careful assembly of all the required features and functionalities which were important to understand pm_{2.5} data. This chapter elaborates on the implementation of the system, data sources, analysis and visualization of the data in a manner that it is easily comprehensible by any user of the application. Correlation techniques used during the implementation further aids the data analysis and correlates to enhance understanding of the impact of pollutant pm_{2.5}. Also described in this chapter is the basic prototype of the application in terms of charts, maps, graphs and tables in order to visualize the data.

4.1. Data Analysis And Visualization

Visualization of data is a way to demonstrate the data in terms of pictures and graphs. Visual representations such as maps and charts help people understand the information more quickly and easily. With the availability and collection of more and more data, there comes a need for visualization software so that this data can be analyzed and presented visually. This helps in discovering how the data is relevant among millions of variables [6].

Data visualization in terms of graphs, charts and maps is always better as compared to going through spreadsheets or reports because they help human brain process information faster. Interactive visualization of data gets a step further in the understanding of data and making it more user-friendly. It digs into the charts and graphs and interactively changes the data that we see and how it is being processed [6].

The availability of the pollution data in category pm_{2.5} was assessed and studied for different countries. However, for the data analysis and visualization purposes, only the data pertaining to different states, cities and counties in the United States were taken into account.

The data was taken from the Emission Inventories from the Technology Transfer Network Clearinghouse for Inventories and Emissions Factors on the U.S Environmental Protection Agency website [7].

The data chosen to be analyzed and visualized was the 2008 and 2011 National Emissions Inventory Data [18]. For the 50 states with their respective counties, criteria air pollutants pm2.5 in 14 major tier categories (or sources of pollution) were selected, analyzed and compared. These 14 tier categories (or sources of pollution) were – petroleum & related industries, fuel combustion and others, solvent utilization, highway vehicles, waste disposal & recycling, miscellaneous, off-highway, other industrial processes, fuel combustion industrial, fuel combustion electrical utility, metals processing, storage & transport, chemical & allied product manufacturing and New Mexico & Utah” respectively. These category tier types as pollutant causes are only pertaining to the pollution data from the U.S. Environmental Protection Agency website.

Since, the data used for analysis and visualization is only for the United States, these category tier types are subject to change as an as data from other countries becomes available. All of the data analysis, comparisons and visualizations have been demonstrated in a series of redirected pages.

The emission data from all the states have been compared on the basis of specific category tiers (or sources of pollution) on the “Visual Representation of Emissions Data” page. For example, emission data as a result of petroleum & related industries in different states have been compared using bar chart where emission (in tons) was plotted against states. Also, this data analysis and comparison has been furthered to the level of counties in a particular state on the “More Visualizations for Emissions data” page. For example, the emission data as a result of

petroleum & related industries in different counties in the state of North Dakota has been visualized and compared. All of this emission data has been plotted in terms of bar charts and geographical maps.

On the “Visual Representation of Emissions Data” page, the data in the bar charts for different states is dynamically displayed on the basis of the category tier (or sources of pollution) selection from the category tier dropdown menu option. This tier selection in the dropdown menu option is also been made to associate with the data populated in a grid view. The emission data in the grid view displays data in terms of states, counties, tier name and emission (in tons). It corresponds to the emissions in accordance with the selected tier category from the category tier (or sources of pollution) dropdown menu option.

When the user clicks on the “Click here” hyperlink, the visualizations for the counties from a state are displayed according to a particular selected state and a corresponding category tier. Upon clicking the “Home” hyperlink, the user is redirected to the Home page. On the More Visualizations for Emissions Data page, when the user clicks the “Visual” hyperlink, the user is redirected to the Visualization Representation of Emissions Data page. On the More Visualizations for Emissions Data page, when the user clicks the “Correlation” hyperlink, the user is redirected to the Correlation of Emissions Data page.

Other kind of data which was chosen to be visually represented was the Air Quality Index data. The data was taken from the U.S. Environmental Protection Agency website and was analyzed and visualized for the years ranging from 1980 to 2013. This data is represented in the form of motion chart which will allow the user to view the trend of the air quality index from 1980 to 2013. User will have the option to toggle between the color codes and the states. This

chart is very much interactive and user friendly and provides a lot of information and comparison of the air quality index. The user can easily access the motion chart through the User Home page.

Data analysis would be incomplete without the inclusion of correlation techniques. In terms of Statistics, the term correlation refers to the relationship between sets of data or any two entities or variables. Correlation also involves statistical dependence on each other. It defines the relationship between variables as to how far the variables tend to vary or tend to be similar or how is the association between them. This can be done in terms of graphs to indicate a positive (meaning highly correlated), negative (meaning least correlated) or a zero relationship (zero relationship meaning no relationship at all).

The correlation technique that is used in this research is the Pearson product-moment correlation coefficient or Pearson correlation coefficient. It measures the strength of the linear association that exists between two variables [20]. It is denoted by 'r' and the formula used to calculate the correlation coefficient is given below. [21].

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

In the formula, x and y are two data sets.

The calculation of the correlation coefficient using this equation models the linear association between those two variables. If $r = 1$, there is a perfect correlation whereas $r = -1$ represents a perfect negative correlation [21]. While a correlation coefficient near ± 1 signifies a near-perfect degree of association between the two variables, a coefficient value near 0 indicates a non-existent or very weak relationship between them [22].

When we are comparing two data sets or two variables, the goal is to determine if there is a relationship between them. To verify whether or not they are correlated, we evaluate the behavior of the variables with respect to each other. If the second variable has a tendency to increase as the first one increases, it implies that there is a positive correlation. If the second variable has a tendency to decrease as the first one increases, it implies that there is a negative correlation. If the second variable neither increases nor decreases as the first one increase, there is no correlation. A positive correlation is indicated by a positive r value, negative correlation is indicated by a negative r value and no correlation is implied by an r value of 0. An r value near 1 or -1 indicates stronger linear correlation [23].

After calculating the r value, it is important that we determine the probability that the r value and the correlation results are valid and have not occurred by chance. This determination is made by first assuming an alpha α value, which is usually 0.05 for checking correlation. An alpha value of 0.05 means that there is a 5 out of 100 chance that the correlation results occurred by chance or that 5 variable sets showed a correlation when there was no correlation between them. Another explanation of this alpha value is that there is a 95 out of 100 chance that correlation is valid and significant. For the purposes of this research work, I have used the alpha value of 0.05 [24].

After the alpha value is determined, the degrees of freedom (df) must be established. To study the correlation,

$$\text{Degrees of freedom (df)} = \text{number of data sets} - 2.$$

For example, if the total number of data sets is 33, the degrees of freedom (df) = 33 - 2 = 31.

After determining the degrees of freedom (df), next step is to determine t-value. The t-value is calculated using the formula below [25].

$$t = \frac{r}{\text{sqrt}[(1-r^2)/(N-2)]}$$

Where N = number of data sets or the size of the sample

r = correlation coefficient

The p-value is then calculated using the following formula [26]:

$$P \text{ Value} = [1 / ((\sqrt{df}) B(1/2, df/2))] \int_{|t|}^{\infty} (1 + x^2/df)^{-(v+1)/2} .dx$$

Where df = degrees of freedom and X = T-value

If the one tailed (or one directional) or the two tailed (or bidirectional or bivariate) probability values, $p < 0.05$, there is a significant statistical correlation. Alternatively stated, the hypothesis we chose with the alpha value of 0.05 is correct and that we are 95% confident that the correlation results are valid and did not occur by chance [27].

4.2. Pollution Analysis System

The working prototype of the PAS will be covered by following through the workflow along with the screenshots. This system only works with the 2008 and 2011 National Emissions Inventory Data and Air Quality Index Data [3] from the U.S. Environmental Protection Agency website [18]. The data downloaded from the website was in the form of .csv files. This system will allow the Admin users to be able to upload any emissions data or the air quality index data in the form of .csv files. This user page where the data can be uploaded will only be available to the Admin users of the website and the normal users will not have access to this webpage. For

now, this page will only accept the .csv files. Future work, however, will include working with other file extensions.

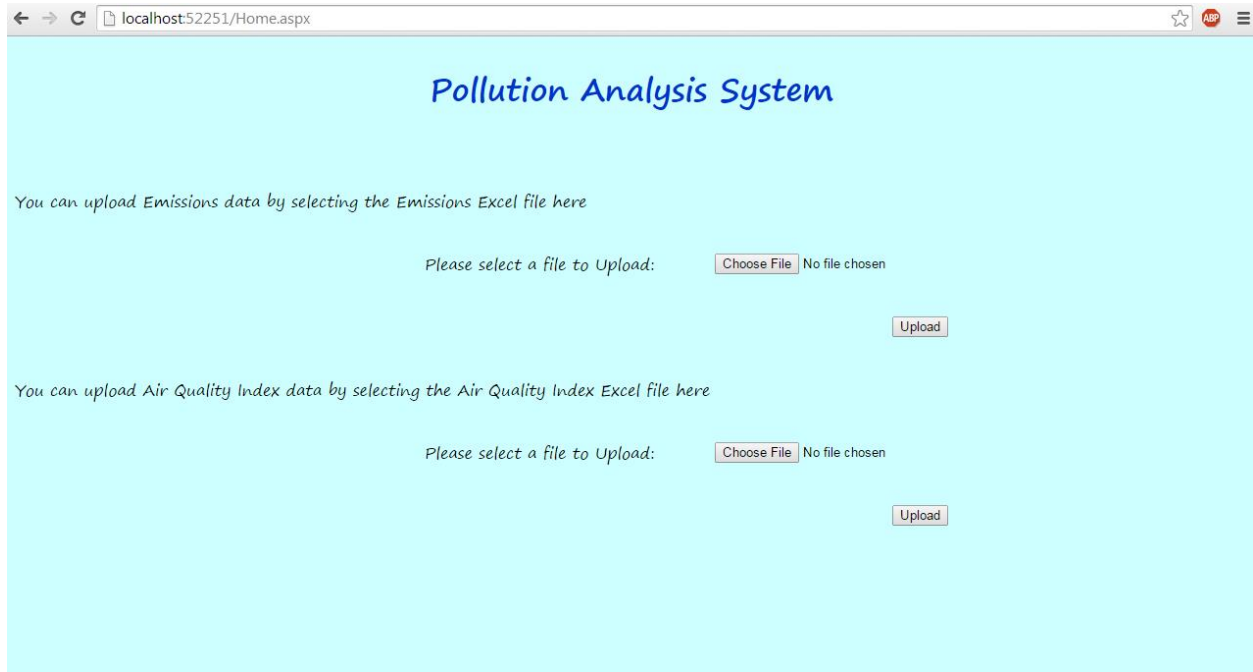


Figure 5. Admin User

By clicking on the first Choose File option button, the user can select an Emissions file whose data needs to be analyzed and visualized. Clicking on the Upload button, the user will be able to upload the Emissions .csv file into the system. The system then uploads this file on the server and into the database with a unique fileId. The contents of the file are stored in the “datafromfile” table in the database.

Upon clicking on the second Choose File option button, the user can select an Air Quality Index file whose data needs to be analyzed and visualized. Clicking on the Upload button, the user will be able to upload the Air Quality Index .csv file into the system. The system then uploads this file on the server and into the database with a unique fileId. The contents of the file are stored in the “airqualityindex” table in the database.

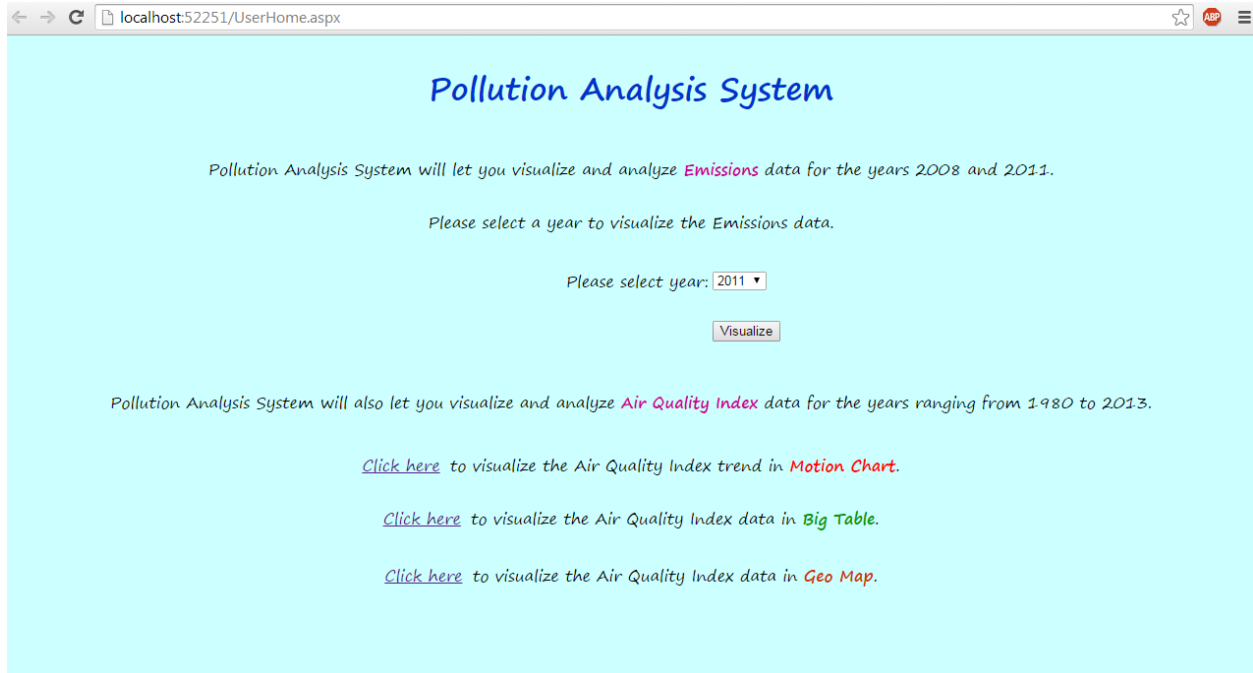


Figure 6. User home

This display is the common application home page for both Admin user and Regular user of the application. Once the user arrives on this home page (figure above), the user can select a year whose data user wishes to visualize and analyze and click “Visualize” button. After the Admin user has uploaded the files in the system, all the files are stored in the database with a unique fileId and has a file number assigned to it. As the Admin user continues to add more data files in the system, the year dropdown list will continue to grow, providing the user with more year options. Right now, there is only two years (2008 and 2011) worth of data which has been uploaded for which the data can be visualized. From this display, the user will also be able to analyze and visualize Air Quality Index data for the years ranging from 1980 to 2013. Different types of visualizations available to the user are Motion Chart, Tabular form and Geo Map. There are three hyperlinks present on this display which will redirect the user to their respective forms of visualizations.

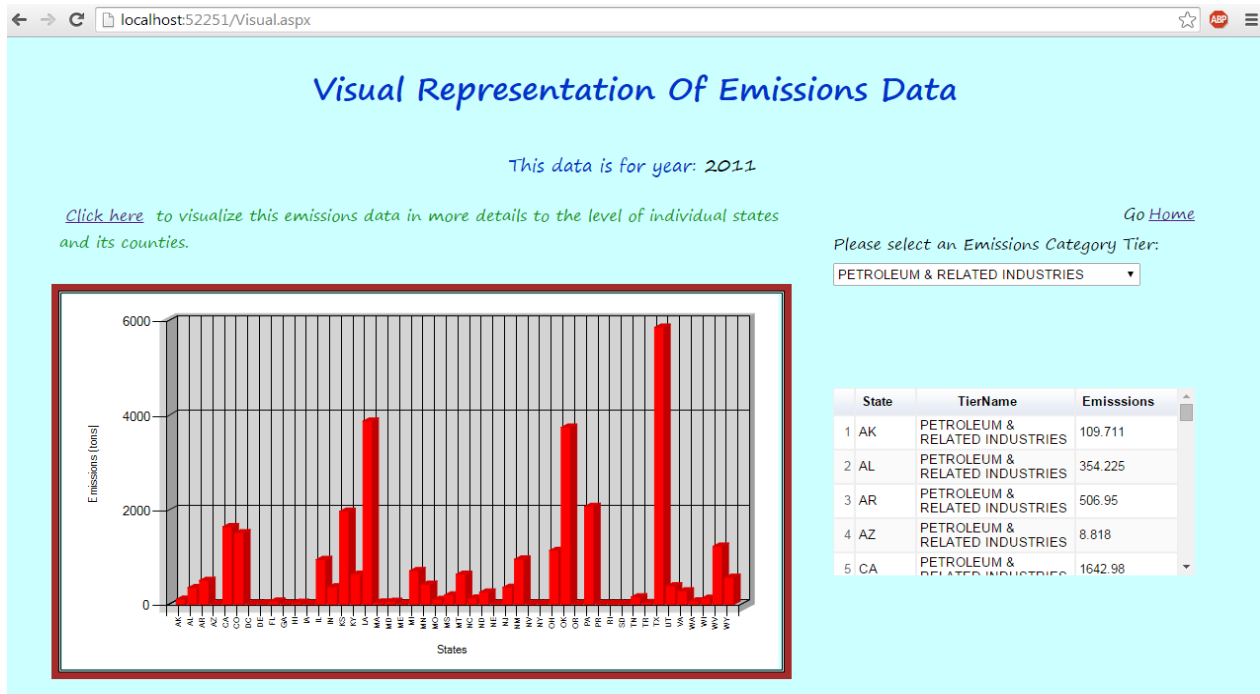


Figure 7. Visual Representation of Emissions Data

After selecting a year and clicking on “Visualize” button, the user is redirected to the “Visual Representation of Emissions Data” display (figure above). On this page, the user can select a pollutant category tier from the category tier (source of pollution) dropdown menu. Upon making a selection, the data corresponding to that particular tier name (pollution source) along with the sum of emission (in tons) for all the counties belonging to a particular state will be populated in the grid view. This grid view is linked to the bar chart.

The data in both the grid view and the bar chart is dynamically populated and modified on the basis of the category tier (source of pollution) selection from the category tier dropdown menu. The columns in the grid view are State, Tier Name and Emissions respectively. The State column contains the state abbreviations. The data in the bar chart is plotted for the Emissions (in tons) against the States. The data in the bar chart here is the sum of the emissions (in tons) of all the counties for a particular state for a particular category tier (source of pollution). For instance, in the screenshot above, state of Texas is reported to have the maximum emission (in tons)

for the category tier Petroleum & Related Industries. This highest bar for the state of Texas corresponds to the sum of the emission (in tons) from all the counties in Texas pertaining to the category tier Petroleum & Related Industries and the fact that this sum is highest as compared to the sum of emission (in tons) from the counties of other states. '0' in the Emissions column by a state corresponds to the fact that no emissions were caused in that particular state for a selected category tier. For instance, the sum of emissions pertaining to category tier type Metals Processing from New Mexico is 0 since no emission was caused by Metals Processing in this state.

Also on this page are the hyperlinks to the "Home" page and for more visualizations of emissions data for if the user would like to visualize the emissions data in more details to the level of individual states and its counties. Clicking on the "Home" hyperlink will redirect the user to the Home page where the user can select a different year whose data the user would like to visualize or if the user would like to visualize the Air Quality Index data. "Click here" hyperlink will redirect the user to the More Visualizations for Emissions data page where the user can select a particular state and the category tier (source of pollution) combination to visualize the emission (in tons) caused in the counties belonging to that particular state as a result of the selected category tier.

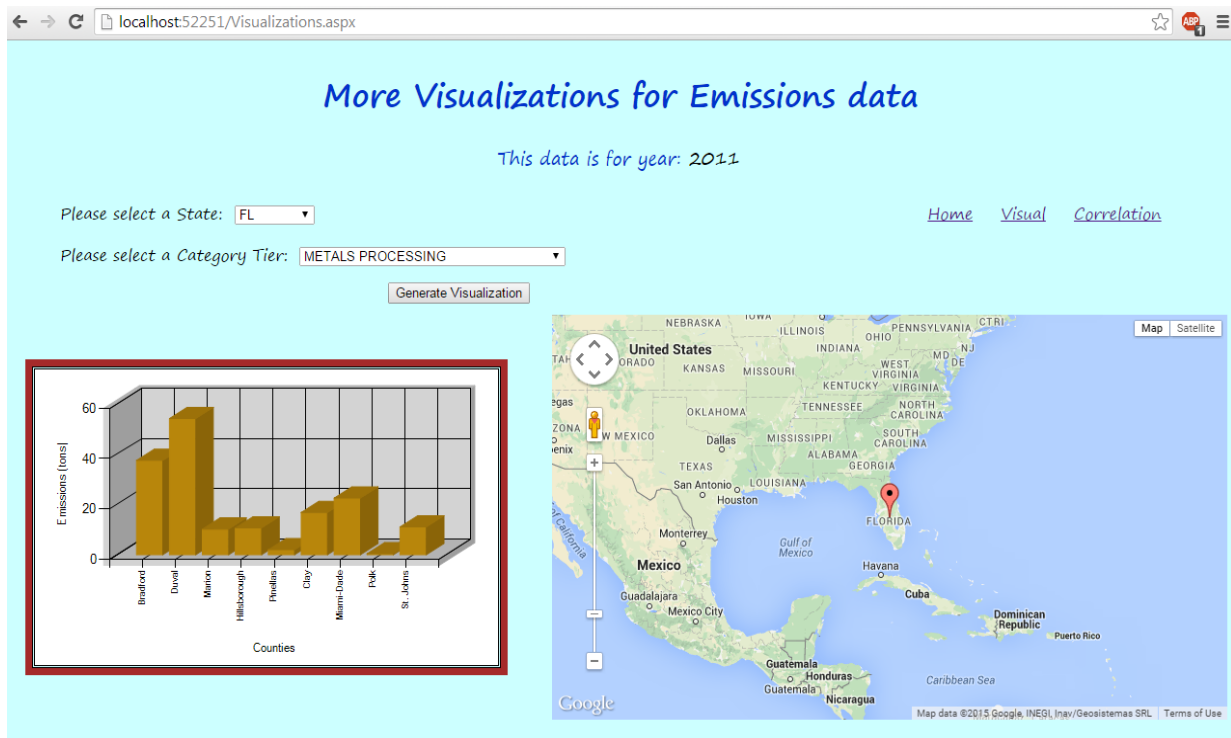


Figure 8. More Visualizations for Emissions data

The user is redirected to this page in the figure above after clicking the “Click here” hyperlink on the “Visual Representation of Emissions Data” page. On this page, the user can select a state from the state dropdown menu and the category tier (source of pollution) from the category tier dropdown menu. The user will have to make a selection from both the dropdown menu options in order to get accurate results. Upon making the appropriate selections and clicking on Generate Visualizations button, a bar chart for the emissions (in tons) against the counties for the selected state and the selected category tier type is generated. Corresponding to this bar chart is the map where the location of the county with the highest emission (in tons) for the selected category tier type is displayed with the marker pin. For instance, the screenshot above displays that the highest emission (in tons) as a result of Metals Processing in the state of Florida in 2011 was reported in Duval County and hence the marker pin in the Duval County in the map.

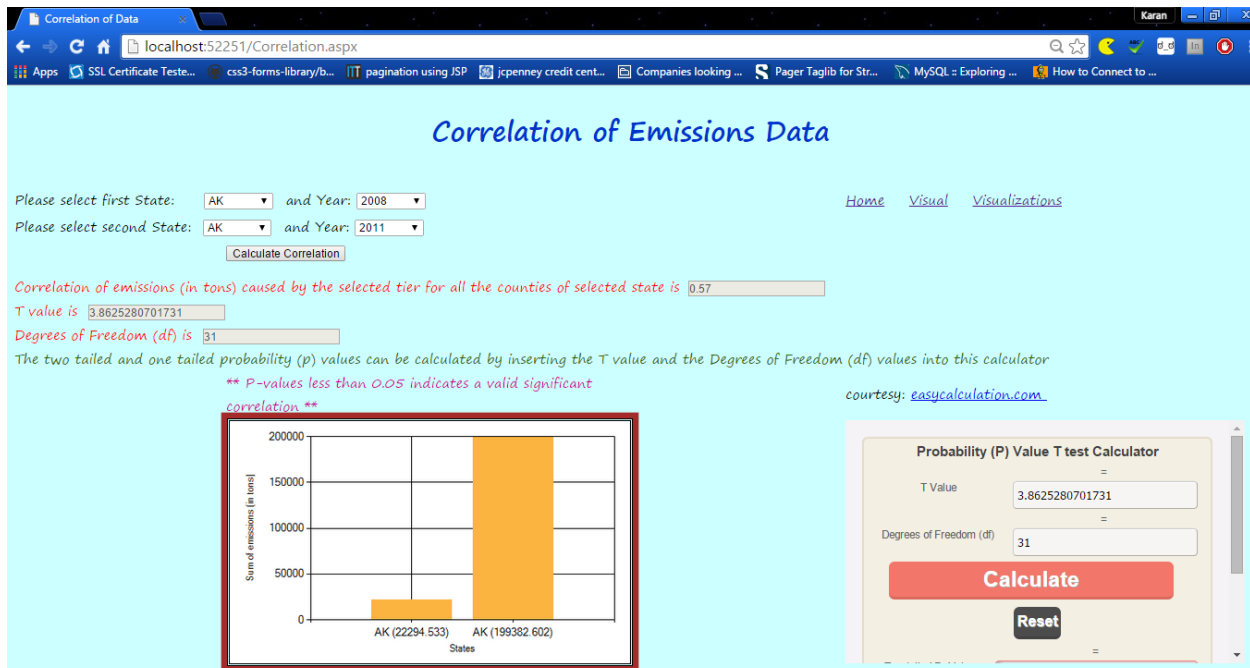


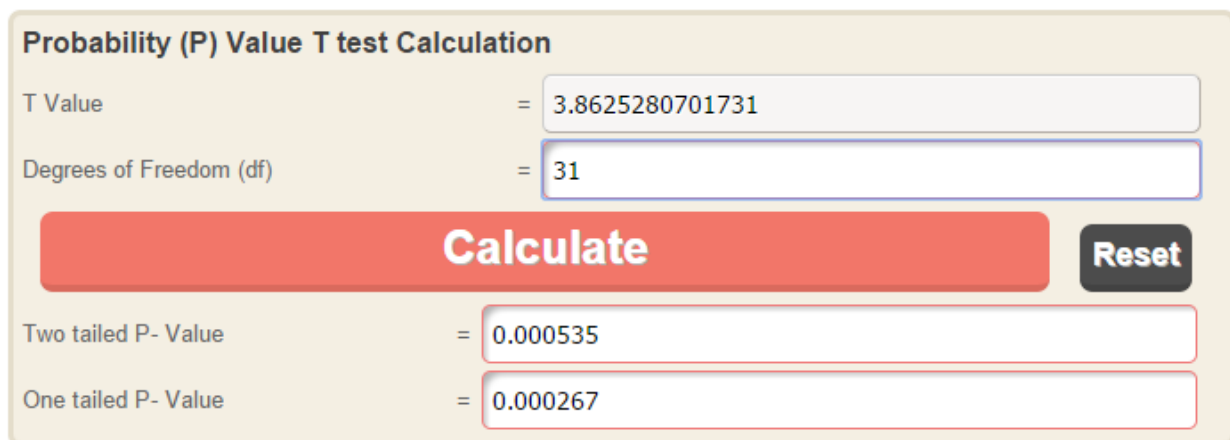
Figure 9. Correlation of Emissions Data

The user is redirected to this page in the figure above after clicking the “Correlation” hyperlink on the “More Visualizations for Emissions data” page. On this page, the user can calculate, view and analyze the correlation between the data from two different states or same states for two different years. The emission data selected here for each state is the total emission (in tons). Total emission here corresponds to the sum of all the emissions caused by all the category tiers in all the counties belonging to a particular state. The user can select a state from the first state dropdown menu and the year from the year dropdown.

Since, the available data is only for years 2008 and 2011 right now, the only year choices user will have would be 2008 and 2011. The user can select a state from the second state dropdown menu and the year from the year dropdown. Upon clicking the “Calculate Correlation” button, the user will be able to view the correlation of the emission data between the two states selected for two different years.

The correlation will only be calculated when both the states are selected. The user will have to make a selection from both the dropdown menu options in order to get accurate results. The correlation can also be calculated for the same state for two different years. Upon making the appropriate selections and clicking on “Calculate Correlation” button, a bar chart for the sum of emission (in tons) against all the counties of the first selected state versus the sum of emission (in tons) for all the counties for the second selected state is generated. For example, in the screenshot above, correlation is calculated between the emission data from all the counties for the same state (Alaska) for years 2008 and 2011.

After the user has clicked the Calculate Correlation button, the correlation coefficient value, degrees of freedom (df) value and t-value is also calculated. The user can plug in these values in the Probability (P) Value T test Calculator and calculate the Two tailed P-value and One tailed P-value. If the P-values come out to be less than 0.05, it indicates that there is a 95% confidence that the correlation is significant. For example, if we calculate the correlation for Alaska for the years 2008 and 2011, the values for correlation coefficient, degrees of freedom (df) and T-value comes out to be 0.57, 3.8625280701731 and 31 respectively.



Probability (P) Value T test Calculation

T Value = 3.8625280701731

Degrees of Freedom (df) = 31

Calculate **Reset**

Two tailed P- Value = 0.000535

One tailed P- Value = 0.000267

Figure 10. Probability (P) Value T test Calculator

When these values are used for the calculation of One tailed and Two tailed P-values, they evaluate to 0.000535 and 0.000267 respectively. Since these values are less than 0.05, the implication is that the correlation is not a fluke and the results are valid and significant.

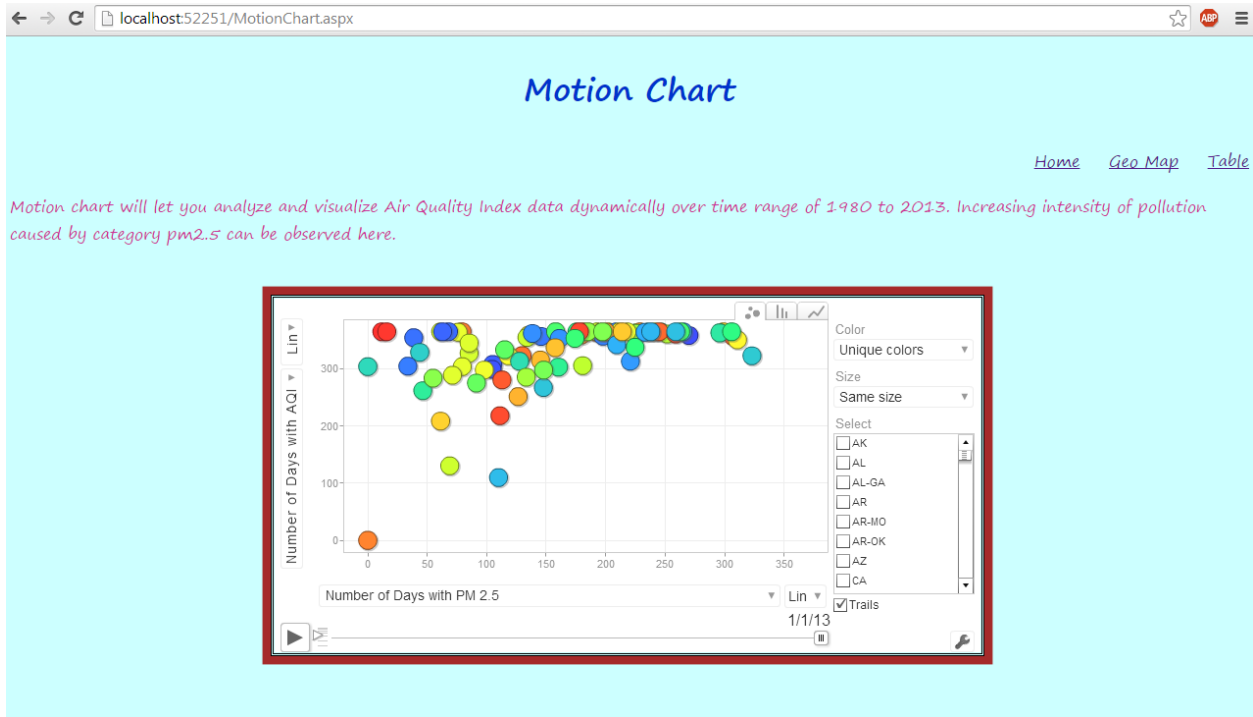


Figure 11. Motion Chart (View 1)

The user can arrive at the page in the figure above by clicking the 'Click Here' link on the Home page. This page displays the motion chart representing the trend of the air quality index of the years ranging from 1980 to 2013. Motion chart can also be used as static chart wherein all the data pertaining to all the states for a year range can be displayed together in just one chart. This chart is drawn against x and y axis where x axis represents Number of Days with pm2.5 and y axis representing number of days with AQI. This motion chart is composed of bubbles where each bubble represents a state. Hovering on the bubble gives information about the air quality index for that particular state for a year. These bubbles are color coded to unique colors to show the difference. The user can also select same color from Color dropdown to the

right of the chart depending on personal preferences. This chart illustration has several settings which can be applied to view the data differently. The other illustrations of this motion chart are

—

- Color settings set to ‘Number of Days with AQI’ meaning the total number of days in a year, for a state, where air quality index was recorded. The balls represent the states and their color represents the intensity where blue being close to 0, yellow being close to 200 which increases and goes all the way to; red being highest.
- Color settings set to ‘Number of Days with PM2.5’ meaning the number of days where pm2.5 was the main pollutant. The balls represent the states and their color represents the intensity where blue being close to 0, yellow being close to 200 which increases and goes all the way to; red being highest.
- Color settings set to ‘Unique colors’. Color settings are set to ‘Unique colors’ meaning each state ball will have a unique color. The size settings are set to ‘Number of Days with AQI’ meaning the total number of days in a year, for a state, where air quality index was recorded. The size of the balls for different states is different and is dependent on the AQI value.
- Color settings set to ‘Unique colors’ meaning each state ball will have a unique color. The size settings are set to ‘Number of Days with PM2.5’ meaning the number of days where pm2.5 was the main pollutant. The size of the balls for different states is different and is dependent on the pm2.5 value.

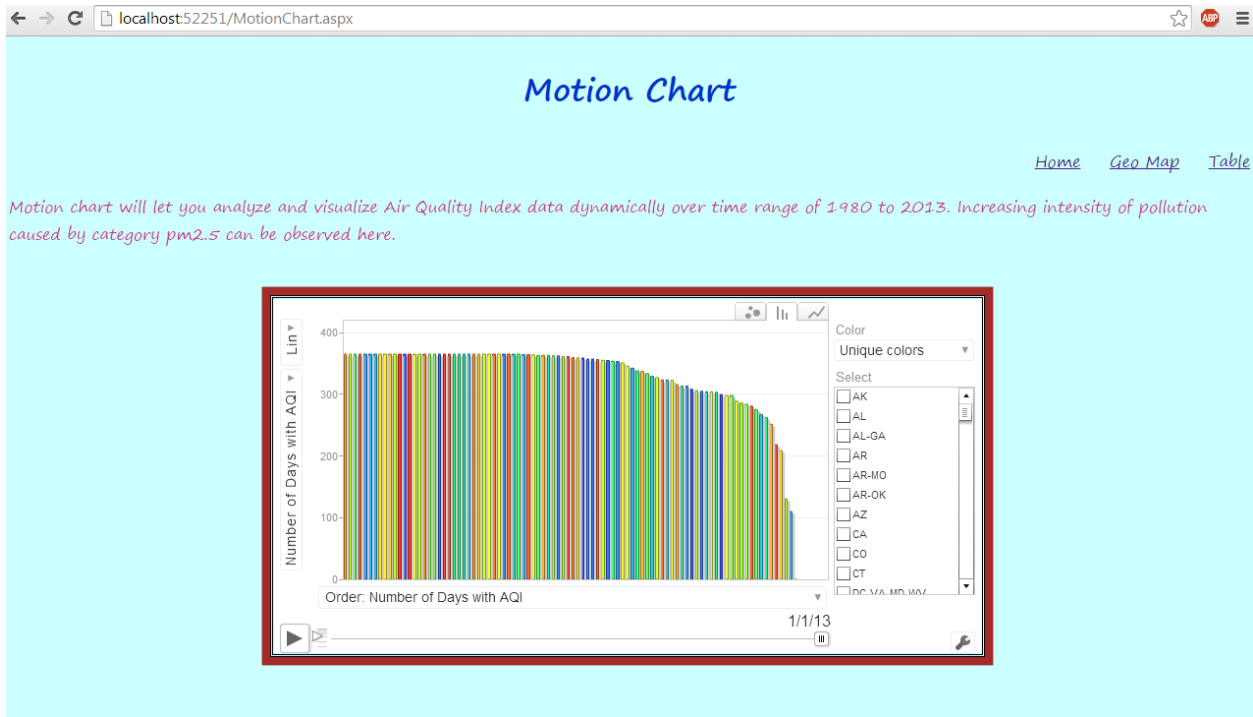


Figure 12. Motion Chart (View 2)

The representation of motion chart in the figure above corresponds to the data trend in terms of bar chart.

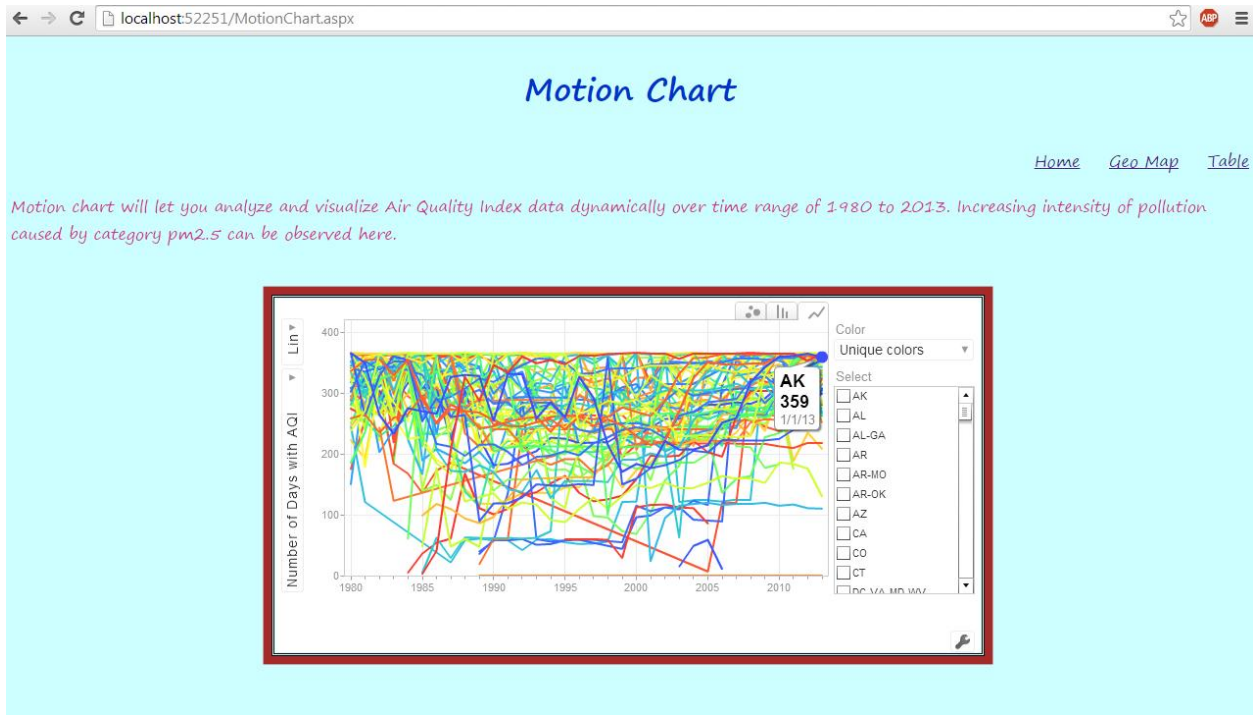


Figure 13. Motion Chart (View 3)

The representation of motion chart in the figure above corresponds to the data trend in terms of line chart where data for Number of Days with AQI for the state of Alaska is displayed along with the year it was recorded for.

Table for Air Quality Index data

Select Year: 1980

[Home](#) [Motion Chart](#) [Geo Map](#)

	City	State	Number Of Days With AQI	Good	Moderate	Unhealthy For Sensitive Groups	Unhealthy	Very Unhealthy	AQIMaximum	Number Of Days PM2.5	
141	Augusta-Waterville	ME	146	119	15		11	1	0	164	0
142	Washington	IN	192	87	74		30	1	0	158	0
143	Safford	AZ	366	89	47		38	192	0	200	0
144	Farmington	NM	366	153	125		77	11	0	200	0
145	Battle Creek	MI	356	356	0		0	0	0	43	0
146	Silver City	NM	366	239	52		55	20	0	200	0
147	Binghamton	NY	366	224	119		22	1	0	151	0
148	Charleston	WV	365	180	134		51	0	0	144	0
149	Green Bay	WI	366	37	64		120	145	0	200	0
150	Honolulu	HI	366	366	0		0	0	0	31	0
151	Vernal	UT	266	265	1		0	0	0	100	0
152	St. Louis	MO-IL	366	2	53		196	103	12	212	0
153	Murray	KY	189	126	41		20	2	0	169	0
154	Shreveport-Bossier City	LA	59	53	3		1	2	0	200	0
155	Berlin	NH-VT	362	46	76		151	89	0	200	0
156	Ocean City	NJ	290	203	46		32	8	1	204	0
157	Charleston	WV	365	180	134		51	0	0	144	0
158	Syracuse	NY	366	138	149		74	5	0	197	0
159	Brunswick	GA	238	235	3		0	0	0	62	0
160	Akron	OH	366	19	123		176	48	0	200	0


Navigation: 1 2 3 4 5 6 7 8 9 10 30 33

Figure 14. Table (View 1)

This display illustrates all the pollution data in category pm2.5 for years ranging from 1980 to 2013 in a tabular form. The user can arrive at this display from the Home page. This kind of visualization offers the user a lot of different ways to visualize data such as sorting, pagination, dynamic display of data upon year selection; and color coded AQI values. The pagination feature of this table will allow the user to leaf through pages of data. This unique feature prevents and reduces data being cluttered on one display and enhances the clarity of data display.

Upon selecting a year from the year dropdown, the user can view data for a particular year. Selection of a year from the year dropdown will dynamically change the data in the table for the user to view it. Also, the values in the Good column are color coded depending upon its intensity for a particular city. For example, the AQI value of 366 for Honolulu, HI is color coded to green to depict that its standing is above and beyond the acceptable or standard of the value of

200 for good AQI value. The table rows are alternately lightly shaded to improve visibility. The data displayed in this table is processed data which was previously processed by the parser when the data files were being uploaded. This table mirrors the data stored in the database which will provide the user with better understanding of the data. This will allow the user to be able to visualize it and analyze it better and act accordingly. Additionally, this table corresponds to; and works alongside the Motion chart in a way that the user will be able to locate the data in this table upon which the Motion chart is based upon.

Also, in this view of the table, data for all the cities is displayed for the year 1980. The cities for which the AQI value is below the normal acceptable or standard of the value of AQI are color coded in red. For example, the AQI value of 145 for Green Bay, WI is color coded to red to depict that its standing is above and beyond the acceptable or standard of the value of 100 for moderate AQI value. Also in this display, can be seen that the pagination has increased when  is clicked and the maximum records a page can display is 20.

localhost:52251/Table.aspx

Table for Air Quality Index data

Select Year: 1985

[Home](#) [Motion Chart](#) [Geo Map](#)

	City	State	Number Of Days With AQI	Good	Moderate	Unhealthy For Sensitive Groups	Unhealthy	Very Unhealthy	AQIMaximum	Number Of Days PM2.5
1	Aberdeen	SD	65	62	3	0	0	0	64	0
2	Aberdeen	WA	353	105	64	142	42	0	200	0
3	Akron	OH	365	78	173	110	4	0	174	0
4	Albany-Lebanon	OR	62	39	23	0	0	0	77	0
5	Albany-Schenectady-Troy	NY	365	259	89	15	2	0	169	0
6	Albuquerque	NM	365	182	173	7	0	3	218	0
7	Allentown-Bethlehem-Easton	PA-NJ	365	179	135	41	10	0	177	0
8	Altoona	PA	365	221	128	14	2	0	161	0
9	Amarillo	TX	40	37	3	0	0	0	59	0
10	Ames	IA	365	280	43	36	6	0	200	0
11	Anchorage	AK	365	228	116	19	2	0	154	0
12	Ann Arbor	MI	192	160	26	6	0	0	145	0
13	Appleton	WI	185	166	17	2	0	0	132	0
14	Asheville	NC	206	178	28	0	0	0	84	0
15	Ashtabula	OH	365	230	91	35	9	0	200	0
16	Athens	TN	323	258	55	9	1	0	200	0
17	Atlanta-Sandy Springs-Marietta	GA	365	156	148	51	10	0	197	0
18	Atlantic City	NJ	365	216	99	36	10	4	212	0
19	Augusta-Richmond County	GA-SC	360	276	61	22	1	0	157	0
20	Augusta-Waterville	ME	304	212	71	17	4	0	195	0

Navigation: 1 2 10 18

Figure 15. Table (View 2)

In this view of the table, data for all the cities is displayed for the year 1985. The user will be able to sort the data in ascending or descending order on each column. The figure above illustrates the data according to the ascending order of cities. Also in this view of the display, is the row highlight feature. The user can select a particular row to highlight it and the row highlight will retain even when the user is leafing through different pages of data. This row shall only un-highlight when the user clicks on it again. For example, the row of data for Atlantic City, NJ is highlighted here in this table view. This row will remain highlighted even when the user is, say, trying to view the data on page 5. This row will only be un-highlighted when the user will click on it again.

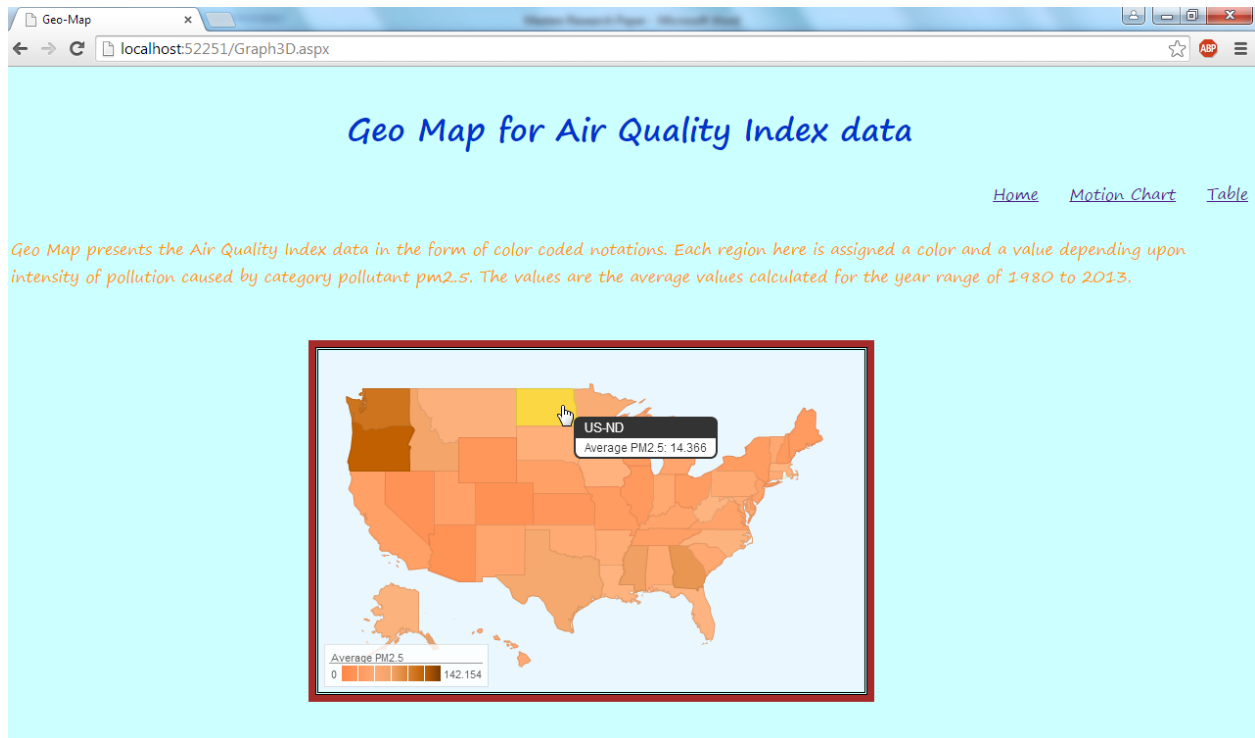


Figure 16. Geo Map

This display illustrates the Geo Map depicting the average of $pm_{2.5}$ values in the years ranging from 1980 to 2013. The color coding for all the states are according to the intensity bar on the bottom left corner which ranges from 0 to the highest value amongst the average values of all the states in the year range of 1980 and 2013 where 0 being the lighter and the maximum value which is 142.154 in this case being the darker shade. Hence, one can roughly sketch an idea that Oregon is the state with maximum emissions consisting of category $pm_{2.5}$ average values followed by Washington.

Also, in this view of the Geo Map in the figure above, the user can view the data for a particular state upon hovering over it. For example, upon hovering over the state of North Dakota, the map of the state is highlighted to a different color and the state code along with the average value for category $pm_{2.5}$ for the year range of 1980 to 2013 is displayed. In this case, the value is approximately 14.366.

CHAPTER 5. PERFORMANCE

Performance of this application has been taken care of at each step while writing the code.

The performance improvement approaches followed while coming up with the structure, design, architecture and coding of this application are as follows –

1. Processing files right away –

The admin page lets the user upload the file. The parser from the admin page takes all the data from the file and processes it. The parse process removes data anomalies such as inconsistent, improper and incompatible data. For example, data containing apostrophe is corrected, so that clean data is retained. After the data is processed, the clean data is saved to the database. This improves performance and reduces the need to further process each time the data is fetched from the database. File upload process is a one-time process but reaching out to the database to fetch appropriate data for the display will happen multiple times over the course of application run. It would have been very inefficient and time consuming had it processed the data for the display every time the database call or connection is made. Processing the file right when it is uploaded seemed an optimal solution.

2. Using List over Arrays –

The States which are displayed in the State dropdown are stored in a list instead of an array. This is because for arrays we have to define a size, but a list can store a large amount of data without having to define the size requirements. List also becomes efficient when we are not aware of the total number of rows worth of data we are about to

store. Similarly, the data for the Emissions Category Tier types (source of pollution) have been stored in the list after they are being fetched from the database.

3. Using data table –

Another performance efficient approach which was taken into account was to reduce the number of database calls. Having a lot of database calls can be expensive and may severely affect the performance of application in terms of making it very slow to load the pages or even breaking the application as a result of query timeouts. In order to improve the efficiency of the application, after the data was fetched from the database, it was stored in the data tables. Data table is a unique inbuilt feature of C# which significantly reduces numerous database calls. This also possesses the ability to store data and perform calculations with the usage of mathematical formulas and notations. Hence, the database won't have to perform strenuous mathematical calculations which might otherwise take a lot of time. A data table also acts like a local hub or a temporary table which can be used to store all the data which has to be displayed or loaded frequently. It makes the application run faster since all the processed data is always available locally and the system won't have to make a call to the database every time the display is loaded.

4. Using more differentiated methods –

While writing the code behind the displays, it was made sure that each and every smallest unit of the application has its dedicated method. Having segregated methods enhances the code flow of the application in terms of readability, functionality and performance on the front end. For example, all the states which compose a state dropdown in any of the displays have a dedicated method in the code which is responsible for fetching all the states from the database for a particular year with the help of SQL queries and then

storing them in a list. This way when the display loads, all it has to do is get all the states from the list, in turn making it a faster process. Besides, since there are a lot of methods and each method has its own dedicated task, it further improves efficiency and reduces processing time for the data load on the display.

CHAPTER 6. TESTING

PAS has been tested thoroughly to ensure proper working of the application as a whole. Each and every module was tested to its entirety. To assure proper working of the application as a result of efficient development; several breakpoints were added to the code to check its validity at each point. To test the completeness and quality of this application these breakpoints were assessed and the results were compared to the actual values from the display. The test cases were designed in Visual Studio 2010 Test Suite using C# as a language. These test cases are represented in tabular form below and are categorized on the basis of the displays for which they were assigned to test.

The following table illustrates the test for the year values on the Pollution Analysis System Home page –

Table 3. Test for year values on User Home page

Display Name – Pollution Analysis System / Home page	
Test 1	To verify if the test passes when all the year values provided are equal and the expected result and the actual result is the years 2011 and 2008 being displayed in the year dropdown menu list.
	Input: Admin user enters Emissions data worth years 2011 and 2008 through the Admin user Home page
	Output: year: 2011, 2008
	Expected Result: year: 2011, 2008

Table 3. Test for year values on User Home page (continued)

```

51 cmd.CommandText = "Select distinct inventoryYear from datafromfile";
52
53 // executing the query
54 rdr = cmd.ExecuteReader();
55
56 // tierName.Add("");
57 while (rdr.Read())
58 {
59     years.Add(rdr.GetString(0));
60 }
61
62 catch (Exception ex)
63 {
64     Console.WriteLine(ex);
65 }
66
67
68 if (connection != null)
69
70     // closing the connection
71     connection.Close();
72
73 return years;
74 }
75
76 protected void V
77 {
78     Session.Add("year", this.yearsDropDownList.SelectedItem.Text);
79     Response.Redirect("Visual.aspx");
80 }
81
82 }

```

Actual Result:
year: 2011, 2008

Test result: Passed

The following table illustrates the test for year value 2011 on Visual Representation of Emissions Data display –

Table 4. Test for year value 2011 on Visual Representation of Emissions Data display

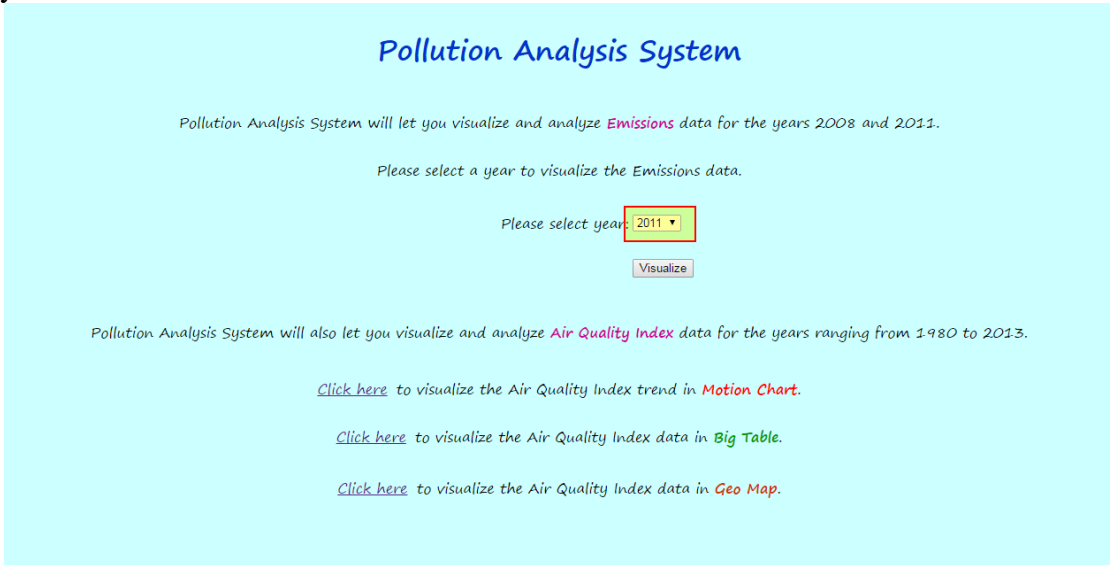
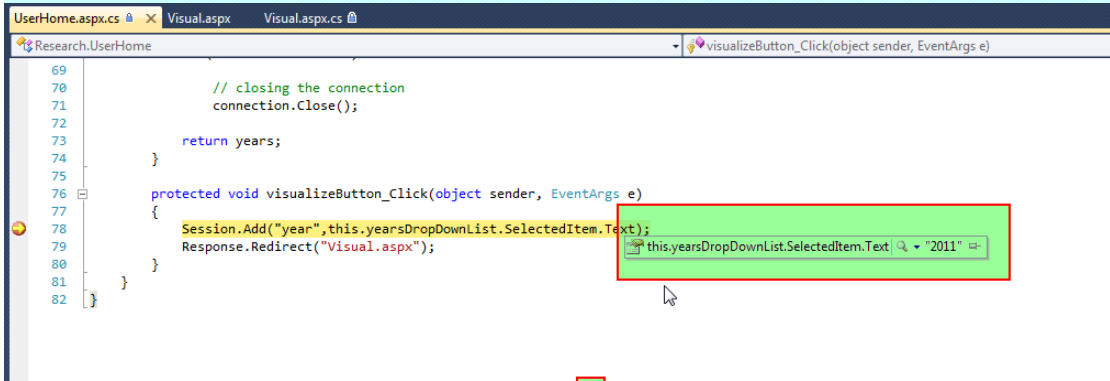
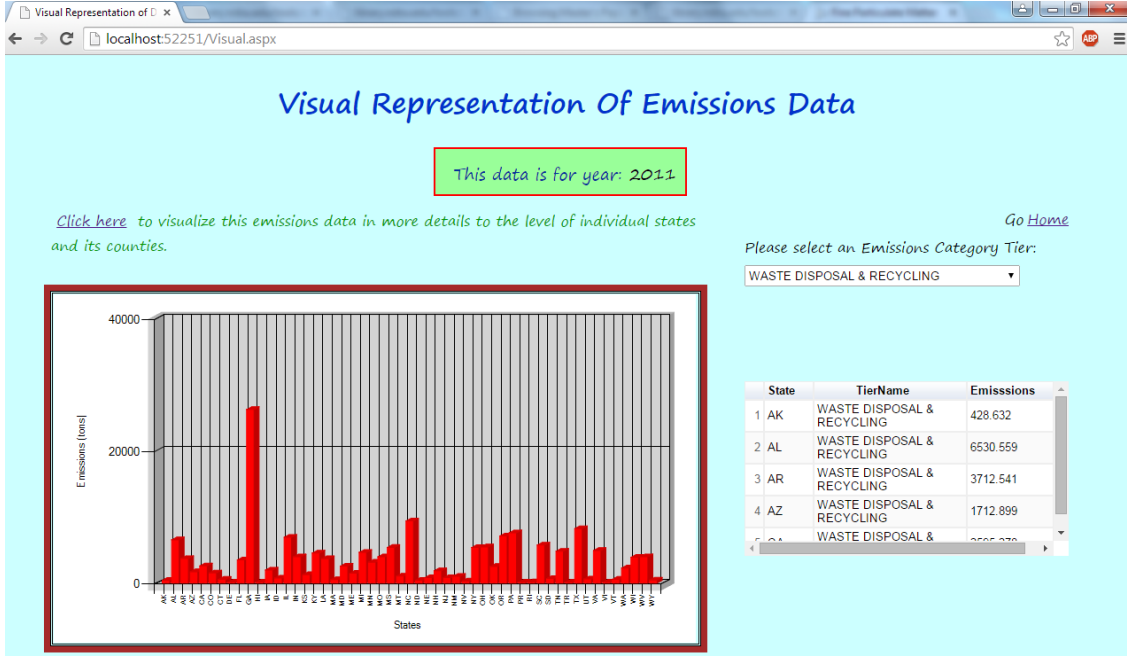
Display Name – Visual Representation of Emissions data	
Test 2	To verify if the test passes when the year value 2011 selected by the user on the Home page is the same as the year value 2011 on the Visual Representation of Emissions data display page and the expected result and the actual result is year value 2011.
	<p>Input: User selected year value 2011 from the year dropdown menu option on the Home page</p> <p>Output: year:2011</p> <p>Expected Result: year:2011</p>  

Table 4. Test for year value 2011 on Visual Representation of Emissions Data display (continued)

	<p>Actual Result: year:2011</p>  <p>Test result: Passed</p>
--	--

The following table illustrates test for year value 2011 on Visual Representation of Emissions Data display –

Table 5. Test for year value 2011 on Visual Representation of Emissions Data display

Display Name – Visual Representation of Emissions data	
Test 3	To verify if the test passes when the year value 2008 selected by the user on the Home page is the same as the year value 2008 on the Visual Representation of Emissions data display page and the expected result and the actual result is year value 2008
	<p>Input: User selected year value 2008 from the year dropdown menu option on the Home page</p> <p>Output: year: 2008</p>

Table 5. Test for year value 2011 on Visual Representation of Emissions Data display (continued)

Expected Result:

year: 2008

The screenshot displays a web browser window at the top and a Visual Studio code editor at the bottom. The browser window shows the 'User Home page' with the URL 'localhost:52251/UserHome.aspx'. The page content includes the title 'Pollution Analysis System' and a message: 'Pollution Analysis System will let you visualize and analyze Emissions data for the years 2008 and 2011. Please select a year to visualize the Emissions data.' Below this is a dropdown menu with '2008' selected and a 'Visualize' button. Further down, there are links for 'Air Quality Index' data: 'Motion Chart', 'Big Table', and 'Geo Map'. The code editor shows the 'visualizeButton_Click' method in 'Visual.aspx.cs', with a red box highlighting the line: 'Session.Add("year", this.yearsDropDownList.SelectedItem.Text);'. A green box highlights the dropdown menu in the browser, showing the selected value '2008'.

Table 5. Test for year value 2011 on Visual Representation of Emissions Data display (continued)

Actual Result: year: 2008

The screenshot shows a web browser window with the URL localhost:52251/Visual.aspx. The page title is "Visual Representation of Emissions Data". A yellow box highlights the text "This data is for year: 2008". Below this, there is a link to visualize data in more detail and a "Go Home" link. A dropdown menu is labeled "Please select an Emissions Category Tier:" and is currently set to "HIGHWAY VEHICLES". To the right of the dropdown is a table with the following data:

State	TierName	Emissions
1 AK	HIGHWAY VEHICLES	1194.372
2 AL	HIGHWAY VEHICLES	3881.073
3 AR	HIGHWAY VEHICLES	2817.723
4 AZ	HIGHWAY VEHICLES	5566.309
5 CA	HIGHWAY VEHICLES	21183.331
6 CO	HIGHWAY VEHICLES	3710.242
7 CT	HIGHWAY VEHICLES	1824.454
8 DC	HIGHWAY VEHICLES	378.066

Test result: Passed

The following table illustrates test for category tier dropdown on Visual Representation of Emissions Data display –

Table 6. Test for category tier dropdown on Visual Representation of Emissions Data display

Display Name – Visual Representation of Emissions data	
Test 4	To verify if the test passes when all the emissions category tier dropdown values being displayed on the Visual Representation of Emissions data display page are same as those stored in the database and the expected result and the actual result display all the emissions category tier dropdown values
	<p>Input:</p> <p>Emissions category tier – WASTE DISPOSAL & RECYCLING METALS PROCESSING OFF-HIGHWAY HIGHWAY VEHICLES SOLVENT UTILIZATION MISCELLANEOUS OTHER INDUSTRIAL PROCESSES FUEL COMB. INDUSTRIAL FUEL COMB. OTHER STORAGE & TRANSPORT FUEL COMB. ELEC. UTIL. CHEMICAL & ALLIED PRODUCT MFG PETROLEUM & RELATED INDUSTRIES</p> <p>Output:</p> <p>Emissions category tier - WASTE DISPOSAL & RECYCLING METALS PROCESSING OFF-HIGHWAY HIGHWAY VEHICLES SOLVENT UTILIZATION MISCELLANEOUS OTHER INDUSTRIAL PROCESSES FUEL COMB. INDUSTRIAL FUEL COMB. OTHER STORAGE & TRANSPORT FUEL COMB. ELEC. UTIL. CHEMICAL & ALLIED PRODUCT MFG PETROLEUM & RELATED INDUSTRIES</p>

Table 6. Test for category tier dropdown on Visual Representation of Emissions Data display (continued)

Expected Result:
 Emissions category tier -
 WASTE DISPOSAL & RECYCLING
 METALS PROCESSING
 OFF-HIGHWAY
 HIGHWAY VEHICLES
 SOLVENT UTILIZATION
 MISCELLANEOUS
 OTHER INDUSTRIAL PROCESSES
 FUEL COMB. INDUSTRIAL
 FUEL COMB. OTHER
 STORAGE & TRANSPORT
 FUEL COMB. ELEC. UTIL.
 CHEMICAL & ALLIED PRODUCT MFG
 PETROLEUM & RELATED INDUSTRIES

The screenshot shows a code editor window for 'Visual.aspx.cs' with a dropdown menu open over the line 'return tierName;'. The dropdown menu, titled 'return tierName; Count = 13', lists the following 13 items:

- [0] "WASTE DISPOSAL & RECYCLING"
- [1] "METALS PROCESSING"
- [2] "OFF-HIGHWAY"
- [3] "HIGHWAY VEHICLES"
- [4] "SOLVENT UTILIZATION"
- [5] "MISCELLANEOUS"
- [6] "OTHER INDUSTRIAL PROCESSES"
- [7] "FUEL COMB. INDUSTRIAL"
- [8] "FUEL COMB. OTHER"
- [9] "STORAGE & TRANSPORT"
- [10] "FUEL COMB. ELEC. UTIL."
- [11] "CHEMICAL & ALLIED PRODUCT MFG"
- [12] "PETROLEUM & RELATED INDUSTRIES"

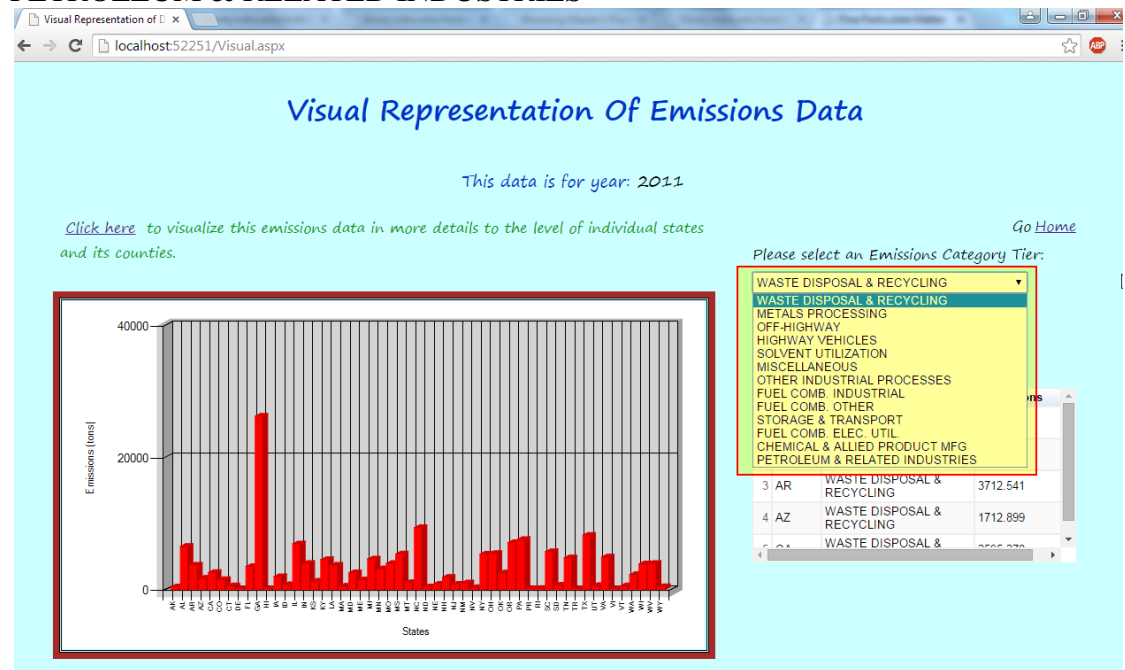
The code in the background includes a try-catch block and a connection closing logic:

```

234     }
235     catch (Exception ex)
236     {
237         Console.WriteLine(ex);
238     }
239
240     if (connection != null)
241
242         // closing the connection
243         connection.Close();
244
245     return tierName;
246 }
247 protected void ...
248 {
249     if (e.Row.R...
250     {
251         if(e.Rc...
252         e.Row.S...
253     }
254 }
255 }
256 }
    
```

Table 6. Test for category tier dropdown on Visual Representation of Emissions Data display (continued)

Actual Result:
 Emissions category tier -
 WASTE DISPOSAL & RECYCLING
 METALS PROCESSING
 OFF-HIGHWAY
 HIGHWAY VEHICLES
 SOLVENT UTILIZATION
 MISCELLANEOUS
 OTHER INDUSTRIAL PROCESSES
 FUEL COMB. INDUSTRIAL
 FUEL COMB. OTHER
 STORAGE & TRANSPORT
 FUEL COMB. ELEC. UTIL.
 CHEMICAL & ALLIED PRODUCT MFG
 PETROLEUM & RELATED INDUSTRIES



Test result: Passed

The following table illustrates test for sum of emissions for all states for a category tier for year 2011 –

Table 7. Test for sum of emissions for all states for a category tier for year 2011

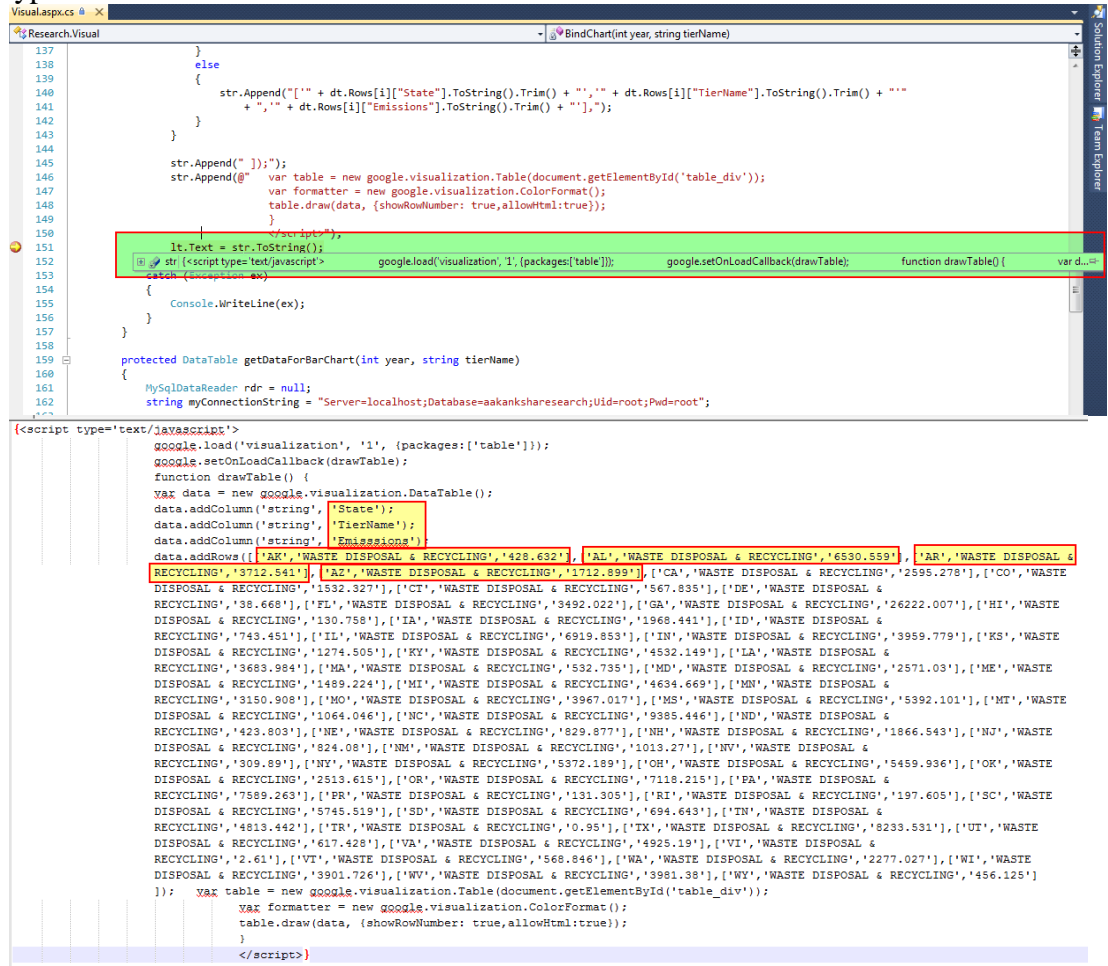
Display Name – Visual Representation of Emissions data	
Test 5	To verify if the test passes when the total sum of all the emissions for a particular state from a particular category tier type as per the calculation from the values in the database matches the total sum value being displayed in the table on the Visual Representation of Emissions data display page and the expected result and the actual result is the total sum of emissions values for a particular state for a particular category tier type for year 2011
	<p>Input: All the Emissions data for the year 2011 for all the states for a particular tier type</p> <p>Output: Total sum of all the emissions for the year 2011 for all the states for a particular tier type</p>
	<p>Expected Result: Total sum of all the emissions for the year 2011 for all the states for a particular tier type</p>  <p>The screenshot shows the following JavaScript code for the drawTable function:</p> <pre> function drawTable() { var data = new google.visualization.DataTable(); data.addColumn('string', 'State'); data.addColumn('string', 'TierName'); data.addColumn('string', 'Emissions'); data.addRow(['AK', 'WASTE DISPOSAL & RECYCLING', '428.632'], ['AL', 'WASTE DISPOSAL & RECYCLING', '6530.559'], ['AR', 'WASTE DISPOSAL & RECYCLING', '3712.541'], ['AZ', 'WASTE DISPOSAL & RECYCLING', '1712.899'], ['CA', 'WASTE DISPOSAL & RECYCLING', '2595.278'], ['CO', 'WASTE DISPOSAL & RECYCLING', '1532.327'], ['CT', 'WASTE DISPOSAL & RECYCLING', '567.835'], ['DE', 'WASTE DISPOSAL & RECYCLING', '38.668'], ['FL', 'WASTE DISPOSAL & RECYCLING', '9492.022'], ['GA', 'WASTE DISPOSAL & RECYCLING', '26222.007'], ['HI', 'WASTE DISPOSAL & RECYCLING', '130.758'], ['IA', 'WASTE DISPOSAL & RECYCLING', '1968.441'], ['ID', 'WASTE DISPOSAL & RECYCLING', '743.451'], ['IL', 'WASTE DISPOSAL & RECYCLING', '6919.853'], ['IN', 'WASTE DISPOSAL & RECYCLING', '3959.779'], ['KS', 'WASTE DISPOSAL & RECYCLING', '1274.505'], ['KY', 'WASTE DISPOSAL & RECYCLING', '4532.149'], ['LA', 'WASTE DISPOSAL & RECYCLING', '3683.984'], ['MA', 'WASTE DISPOSAL & RECYCLING', '532.735'], ['MD', 'WASTE DISPOSAL & RECYCLING', '2571.03'], ['ME', 'WASTE DISPOSAL & RECYCLING', '1489.224'], ['MI', 'WASTE DISPOSAL & RECYCLING', '4634.669'], ['MN', 'WASTE DISPOSAL & RECYCLING', '3150.908'], ['MO', 'WASTE DISPOSAL & RECYCLING', '3967.017'], ['MS', 'WASTE DISPOSAL & RECYCLING', '5392.101'], ['MT', 'WASTE DISPOSAL & RECYCLING', '1064.046'], ['NC', 'WASTE DISPOSAL & RECYCLING', '9385.446'], ['ND', 'WASTE DISPOSAL & RECYCLING', '423.803'], ['NE', 'WASTE DISPOSAL & RECYCLING', '829.877'], ['NH', 'WASTE DISPOSAL & RECYCLING', '1866.543'], ['NJ', 'WASTE DISPOSAL & RECYCLING', '824.08'], ['NM', 'WASTE DISPOSAL & RECYCLING', '1013.27'], ['NV', 'WASTE DISPOSAL & RECYCLING', '309.89'], ['NY', 'WASTE DISPOSAL & RECYCLING', '5372.189'], ['OH', 'WASTE DISPOSAL & RECYCLING', '5459.936'], ['OK', 'WASTE DISPOSAL & RECYCLING', '2513.615'], ['OR', 'WASTE DISPOSAL & RECYCLING', '7118.215'], ['PA', 'WASTE DISPOSAL & RECYCLING', '7589.263'], ['PR', 'WASTE DISPOSAL & RECYCLING', '131.305'], ['RI', 'WASTE DISPOSAL & RECYCLING', '197.605'], ['SC', 'WASTE DISPOSAL & RECYCLING', '5745.519'], ['SD', 'WASTE DISPOSAL & RECYCLING', '694.643'], ['TN', 'WASTE DISPOSAL & RECYCLING', '4813.442'], ['TX', 'WASTE DISPOSAL & RECYCLING', '0.95'], ['UT', 'WASTE DISPOSAL & RECYCLING', '8233.531'], ['VT', 'WASTE DISPOSAL & RECYCLING', '617.428'], ['VA', 'WASTE DISPOSAL & RECYCLING', '4925.19'], ['VI', 'WASTE DISPOSAL & RECYCLING', '2.61'], ['VT', 'WASTE DISPOSAL & RECYCLING', '568.846'], ['WA', 'WASTE DISPOSAL & RECYCLING', '2277.027'], ['WI', 'WASTE DISPOSAL & RECYCLING', '3901.726'], ['WV', 'WASTE DISPOSAL & RECYCLING', '3981.38'], ['WY', 'WASTE DISPOSAL & RECYCLING', '456.125']]; var table = new google.visualization.Table(document.getElementById('table_div')); var formatter = new google.visualization.ColorFormat(); table.draw(data, {showRowNumber: true, allowHtml: true}); } </pre>
	<p>Actual Result: Total sum of all the emissions for the year 2011 for all the states for a particular tier type</p>

Table 7. Test for sum of emissions for all states for a category tier for year 2011 (continued)

Visual Representation Of Emissions Data

This data is for year: 2011

[Click here](#) to visualize this emissions data in more details to the level of individual states and its counties.

[Go Home](#)

Please select an Emissions Category Tier:

State	TierName	Emissions
1 AK	WASTE DISPOSAL & RECYCLING	428.632
2 AL	WASTE DISPOSAL & RECYCLING	6530.559
3 AR	WASTE DISPOSAL & RECYCLING	3712.541
4 AZ	WASTE DISPOSAL & RECYCLING	1712.899

Test result: Passed

The following table illustrates test for sum of emissions for all states for a category tier for year 2008 –

Table 8. Test for sum of emissions for all states for a category tier for year 2008

Display Name – Visual Representation of Emissions data	
Test 6	To verify if the test passes when the total sum of all the emissions for a particular state from a particular category tier type as per the calculation from the values in the database matches the total sum value being displayed in the table on the Visual Representation of Emissions data display page and the expected result and the actual result is the total sum of emissions values for a particular state for a particular category tier type for year 2008
	Input: All the Emissions data for the year 2008 for all the states for a particular tier type Output: Total sum of all the emissions for the year 2008 for all the states for a particular tier type
	Expected Result: Total sum of all the emissions for the year 2008 for all the states for a particular tier type

Table 8. Test for sum of emissions for all states for a category tier for year 2008 (continued)

```

133         if (i == count)
134         {
135             str.Append("[" + dt.Rows[i]["State"].ToString().Trim() + "," + dt.Rows[i]["TierName"].ToString().Trim() + ""
136                 + "," + dt.Rows[i]["Emissions"].ToString().Trim() + ""]);
137         }
138         else
139         {
140             str.Append("[" + dt.Rows[i]["State"].ToString().Trim() + "," + dt.Rows[i]["TierName"].ToString().Trim() + ""
141                 + "," + dt.Rows[i]["Emissions"].ToString().Trim() + ""]);
142         }
143     }
144
145     str.Append("]");
146     var table = new google.visualization.Table(document.getElementById('table_div'));
147     var formatter = new google.visualization.ColorFormat();
148     table.draw(data, {showRowNumber: true, allowHtml: true});
149 }
150
151 </script>
152 <script type="text/javascript"> google.load('visualization', '1', {packages:['table']}); google.setOnLoadCallback(drawTable); function drawTable() {
153     catch (Exception ex)
154     {
155         Console.WriteLine(ex);
156     }
157 }
158
159 protected DataTable getDataForBarChart(int year, string tierName)
160 {
161     MySqlDataReader rdr = null;
162     string myConnectionString = "Server=localhost;Database=aaanksharesearch;Uid=root;Pwd=root";
163
164 </script type="text/javascript">
165     google.load('visualization', '1', {packages:['table']});
166     google.setOnLoadCallback(drawTable);
167     function drawTable() {
168         var data = new google.visualization.DataTable();
169         data.addColumn('string', 'State');
170         data.addColumn('string', 'TierName');
171         data.addColumn('string', 'Emissions');
172         data.addRows([["AK", "HIGHWAY VEHICLES", "1194.372"], ["AL", "HIGHWAY VEHICLES", "3881.073"], ["AR", "HIGHWAY
173     VEHICLES", "2817.723"], ["AZ", "HIGHWAY VEHICLES", "5566.309"], ["CA", "HIGHWAY VEHICLES", "21183.331"], ["CO", "HIGHWAY
174     VEHICLES", "3710.242"], ["CT", "HIGHWAY VEHICLES", "1824.454"], ["DC", "HIGHWAY VEHICLES", "378.099"], ["DE", "HIGHWAY
175     VEHICLES", "498.328"], ["FL", "HIGHWAY VEHICLES", "14103.106"], ["GA", "HIGHWAY VEHICLES", "11887.548"], ["HI", "HIGHWAY
176     VEHICLES", "821.098"], ["IA", "HIGHWAY VEHICLES", "3191.81"], ["ID", "HIGHWAY VEHICLES", "797.868"], ["IL", "HIGHWAY
177     VEHICLES", "10676.686"], ["IN", "HIGHWAY VEHICLES", "6609.087"], ["KS", "HIGHWAY VEHICLES", "2736.653"], ["KY", "HIGHWAY
178     VEHICLES", "4281.298"], ["LA", "HIGHWAY VEHICLES", "3422.043"], ["MA", "HIGHWAY VEHICLES", "3941.016"], ["MD", "HIGHWAY
179     VEHICLES", "3618.809"], ["ME", "HIGHWAY VEHICLES", "1386.578"], ["MI", "HIGHWAY VEHICLES", "10897.039"], ["MN", "HIGHWAY
180     VEHICLES", "7383.87"], ["MO", "HIGHWAY VEHICLES", "7010.722"], ["MS", "HIGHWAY VEHICLES", "2679.561"], ["MT", "HIGHWAY
181     VEHICLES", "1231.41"], ["NC", "HIGHWAY VEHICLES", "7895.427"], ["ND", "HIGHWAY VEHICLES", "889.948"], ["NE", "HIGHWAY
182     VEHICLES", "2276.376"], ["NH", "HIGHWAY VEHICLES", "1145.458"], ["NJ", "HIGHWAY VEHICLES", "4863.235"], ["NM", "HIGHWAY
183     VEHICLES", "2332.376"], ["NV", "HIGHWAY VEHICLES", "1221.75"], ["NY", "HIGHWAY VEHICLES", "8691.643"], ["OH", "HIGHWAY
184     VEHICLES", "10873.542"], ["OK", "HIGHWAY VEHICLES", "4352.701"], ["OR", "HIGHWAY VEHICLES", "3466.577"], ["PA", "HIGHWAY
185     VEHICLES", "7758.968"], ["PR", "HIGHWAY VEHICLES", "2832.77"], ["RI", "HIGHWAY VEHICLES", "462.47"], ["SC", "HIGHWAY
186     VEHICLES", "3404.069"], ["SD", "HIGHWAY VEHICLES", "964.234"], ["TN", "HIGHWAY VEHICLES", "6724.292"], ["TX", "HIGHWAY
187     VEHICLES", "21816.932"], ["UT", "HIGHWAY VEHICLES", "2539.483"], ["VA", "HIGHWAY VEHICLES", "5552.881"], ["VI", "HIGHWAY
188     VEHICLES", "20.952"], ["VT", "HIGHWAY VEHICLES", "477.092"], ["WA", "HIGHWAY VEHICLES", "7051.802"], ["WI", "HIGHWAY
189     VEHICLES", "4589.454"], ["WV", "HIGHWAY VEHICLES", "1565.304"], ["WY", "HIGHWAY VEHICLES", "1100.472"] ]); var table = new
190     google.visualization.Table(document.getElementById('table_div'));
191     var formatter = new google.visualization.ColorFormat();
192     table.draw(data, {showRowNumber: true, allowHtml: true});
193 }
194 </script>

```

Actual Result:
 Total sum of all the emissions for the year 2008 for all the states for a particular tier type

57

Table 8. Test for sum of emissions for all states for a category tier for year 2008 (continued)

Visual Representation Of Emissions Data

This data is for year: 2008

Click [here](#) to visualize this emissions data in more details to the level of individual states and its counties.

Go [Home](#)

Please select an Emissions Category Tier:

HIGHWAY VEHICLES ▼

State	TierName	Emissions
1 AK	HIGHWAY VEHICLES	1194.372
2 AL	HIGHWAY VEHICLES	3881.073
3 AR	HIGHWAY VEHICLES	2817.723
4 AZ	HIGHWAY VEHICLES	5566.309
5 CA	HIGHWAY VEHICLES	21183.331
6 CO	HIGHWAY VEHICLES	3710.242
7 CT	HIGHWAY VEHICLES	1824.454
8 DC	HIGHWAY VEHICLES	378.099

Test result: Passed

The following table illustrates test for selected year value 2011 on More Visualizations for Emissions data –

Table 9. Test for selected year value 2011 on More Visualizations for Emissions data

Display Name – More Visualizations for Emissions data	
Test 7	To verify if the test passes when the year value 2011 selected by the user on the Home page is the same as the year value 2011 on the Visual Representation of Emissions data display page and the expected result and the actual result is year value 2011
	Input: User selected year value 2011 from the year dropdown menu option on the Home page
	Output: year: 2011
	Expected Result: year: 2011

Table 9. Test for selected year value 2011 on More Visualizations for Emissions data (continued)

The screenshot shows a web application titled "Pollution Analysis System". The page content includes:

- Header: "Pollution Analysis System"
- Text: "Pollution Analysis System will let you visualize and analyze Emissions data for the years 2008 and 2011."
- Text: "Please select a year to visualize the Emissions data."
- Form: "Please select year:" followed by a dropdown menu with "2011" selected and a "Visualize" button.
- Text: "Pollution Analysis System will also let you visualize and analyze Air Quality Index data for the years ranging from 1980 to 2013."
- Links: Three links with text: "Click here to visualize the Air Quality Index trend in Motion Chart.", "Click here to visualize the Air Quality Index data in Big Table.", and "Click here to visualize the Air Quality Index data in Geo Map."

Below the web page is a code editor window showing the implementation of the "Visualize" button click event:

```
69  
70 // closing the connection  
71 connection.Close();  
72  
73 return years;  
74 }  
75  
76 protected void visualizeButton_Click(object sender, EventArgs e)  
77 {  
78     Session.Add("year", this.yearsDropDownList.SelectedItem.Text);  
79     Response.Redirect("Visual.aspx");  
80 }  
81 }  
82 }
```

The code editor also shows a debugger window with the following expression: `this.yearsDropDownList.SelectedItem.Text` with a value of `"2011"`.

Actual Result:
year: 2011

Table 10. Test for select year value 2008 on More Visualizations for Emissions data (continued)

The screenshot displays a web browser window at localhost:52251/UserHome.aspx. The page title is "Pollution Analysis System". The main content area has a light blue background and contains the following text:

Pollution Analysis System will let you visualize and analyze **Emissions** data for the years 2008 and 2011.

Please select a year to visualize the Emissions data.

Please select year:

Pollution Analysis System will also let you visualize and analyze **Air Quality Index** data for the years ranging from 1980 to 2013.

[Click here](#) to visualize the Air Quality Index trend in **Motion Chart**.

[Click here](#) to visualize the Air Quality Index data in **Big Table**.

[Click here](#) to visualize the Air Quality Index data in **Geo Map**.

Below the browser window, the Visual Studio code editor shows the C# code for the `visualizeButton_Click` event handler in `Visual.aspx.cs`. The code is as follows:

```
71 connection.Close();
72
73 return years;
74 }
75
76 protected void visualizeButton_Click(object sender, EventArgs e)
77 {
78     Session.Add("year", this.yearsDropDownList.SelectedItem.Text);
79     Response.Redirect("Visual.aspx");
80 }
81 }
82 }
```

The value "2008" is highlighted in a red box in the code editor, corresponding to the value in the browser's dropdown menu.

Actual Result:
year : 2008

Table 10. Test for select year value 2008 on More Visualizations for Emissions data (continued)

	<div style="text-align: center;"> <h3 style="color: #4F81BD;">More Visualizations for Emissions data</h3> <div style="border: 1px solid #FFD700; padding: 2px; display: inline-block; margin: 5px auto;">This data is for year: 2008</div> <p>Please select a State: <input type="text" value="AK"/></p> <p>Please select a Category Tier: <input type="text" value="HIGHWAY VEHICLES"/></p> <p><input type="button" value="Generate Visualization"/></p> </div> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="border: 2px solid #8B0000; padding: 5px;"> </div> <div> </div> </div> <p style="margin-top: 10px;">Test result: Passed</p>
--	---

The following table illustrates test for State dropdown –

Table 11. Test for State dropdown

Display Name – More Visualizations for Emissions data	
Test 9	To verify if the test passes when all the states are being correctly displayed in the State dropdown menu
	<p>Input: States – AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, PR, RI, SC, SD, TN, TR, TX, UT, VA, VI, VT, WA, WI, WV, WY</p> <p>Output: States - AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, PR, RI, SC, SD, TN, TR, TX, UT, VA, VI, VT, WA, WI, WV, WY</p>
	<p>Expected Result: States - AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, PR, RI, SC, SD, TN, TR, TX, UT, VA, VI, VT, WA, WI, WV, WY</p>

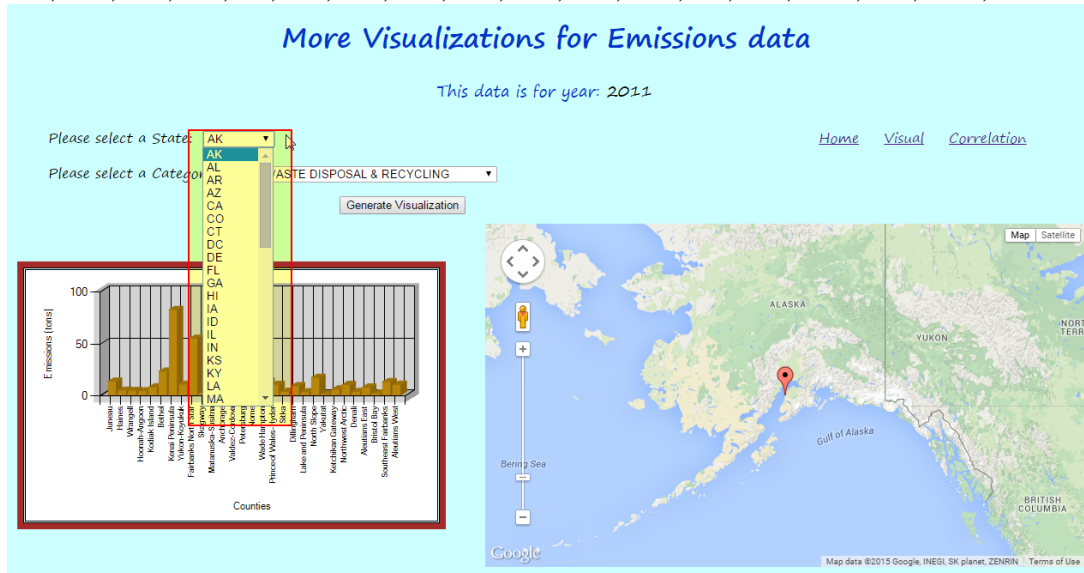
Table 11. Test for State dropdown (continued)

```

Visualizations.aspx.cs
Research.Visualizations
getListOfUsStates(int year)
187 cmd = connection.CreateCommand();
188
189 // to get all the distinct states
190 cmd.CommandText = "Select distinct state from datafromfile where inventoryYear=@year order by state";
191
192 cmd.Parameters.AddWithValue("@year", year);
193
194 // executing the query
195 rdr = cmd.ExecuteReader();
196
197 while (rdr.Read())
198 {
199     string state = rdr.GetString(0);
200 }
201 }
202 catch (Exception ex)
203 {
204     Console.WriteLine(ex.Message);
205 }
206
207 if (connection.State == ConnectionState.Open)
208     connection.Close();
209
210 return state;
211
212 }
213
214 protected string getCountyWithMaxEmission(int year, string state, string tierName)
215 {
216     string countyName = string.Empty;
217     MySqlDataReader rdr = null;
218
219 }
    
```

Actual Result:

States - AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, PR, RI, SC, SD, TN, TR, TX, UT, VA, VI, VT, WA, WI, WV, WY



Test result: Passed

This test illustrates test for Category Tier dropdown on More Emissions of Emissions

Data display –

Table 12. Test for Category Tier dropdown on More Emissions of Emissions Data display

Display Name – More Visualizations for Emissions data	
Test 10	To verify if the test passes when all the emissions category tier dropdown values being displayed on the More Visualizations for Emissions data display page are same as those stored in the database and the expected result and the actual result display all the emissions category tier dropdown values
	<p>Input:</p> <p>Emissions category tier – WASTE DISPOSAL & RECYCLING METALS PROCESSING OFF-HIGHWAY HIGHWAY VEHICLES SOLVENT UTILIZATION MISCELLANEOUS OTHER INDUSTRIAL PROCESSES FUEL COMB. INDUSTRIAL FUEL COMB. OTHER STORAGE & TRANSPORT FUEL COMB. ELEC. UTIL. CHEMICAL & ALLIED PRODUCT MFG PETROLEUM & RELATED INDUSTRIES</p> <p>Output:</p> <p>Emissions category tier – WASTE DISPOSAL & RECYCLING METALS PROCESSING OFF-HIGHWAY HIGHWAY VEHICLES SOLVENT UTILIZATION MISCELLANEOUS OTHER INDUSTRIAL PROCESSES FUEL COMB. INDUSTRIAL FUEL COMB. OTHER STORAGE & TRANSPORT FUEL COMB. ELEC. UTIL. CHEMICAL & ALLIED PRODUCT MFG PETROLEUM & RELATED INDUSTRIES</p>

Table 12. Test for Category Tier dropdown on More Emissions of Emissions Data display (continued)

Expected Result:
 Emissions category tier –
 WASTE DISPOSAL & RECYCLING
 METALS PROCESSING
 OFF-HIGHWAY
 HIGHWAY VEHICLES
 SOLVENT UTILIZATION
 MISCELLANEOUS
 OTHER INDUSTRIAL PROCESSES
 FUEL COMB. INDUSTRIAL
 FUEL COMB. OTHER
 STORAGE & TRANSPORT
 FUEL COMB. ELEC. UTIL.
 CHEMICAL & ALLIED PRODUCT MFG
 PETROLEUM & RELATED INDUSTRIES

```

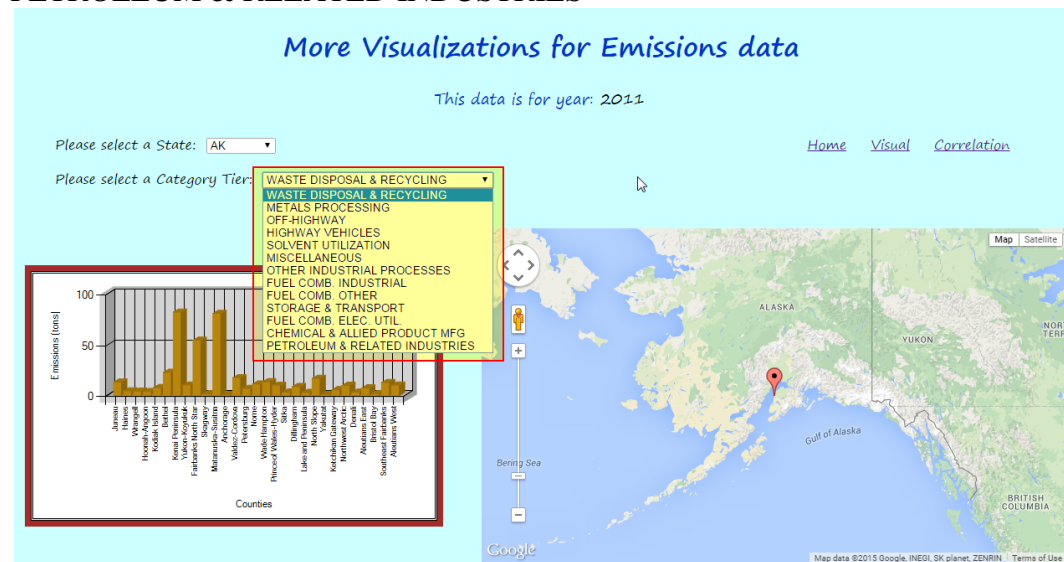
138 cmd = connection.CreateCommand();
139
140 // query for all the tier names for the selected year
141 cmd.CommandText = "Select distinct tierName from datafromfile where inventoryYear=@year";
142
143 cmd.Parameters.AddWithValue("@year", year);
144
145 // executing the query
146 rdr = cmd.ExecuteReader();
147
148 //tierName.Add("");
149 while (rdr.Read())
150 {
151     //string tName = rdr.GetString(0).Trim(' ', '\t', '\n', '\r', '\f', '\n', '');
152     tierName.Add(rdr.GetString(0));
153 }
154 catch (Exception ex)
155 {
156     Console.WriteLine(ex.Message);
157 }
158
159 if (connection.State == ConnectionState.Open)
160 {
161     // closing
162     connection.Close();
163 }
164
165 return tierName;
166 }
167
168 protected List<string> getListOfUsStates(int year)
169 {
170     List<string> states = new List<string>();
    
```

The screenshot shows a dropdown menu with the following items:

- [0] WASTE DISPOSAL & RECYCLING
- [1] METALS PROCESSING
- [2] OFF-HIGHWAY
- [3] HIGHWAY VEHICLES
- [4] SOLVENT UTILIZATION
- [5] MISCELLANEOUS
- [6] OTHER INDUSTRIAL PROCESSES
- [7] FUEL COMB. INDUSTRIAL
- [8] FUEL COMB. OTHER
- [9] STORAGE & TRANSPORT
- [10] FUEL COMB. ELEC. UTIL.
- [11] CHEMICAL & ALLIED PRODUCT MFG
- [12] PETROLEUM & RELATED INDUSTRIES

Table 12. Test for Category Tier dropdown on More Emissions of Emissions Data display (continued)

Actual Result:
 Emissions category tier –
 WASTE DISPOSAL & RECYCLING
 METALS PROCESSING
 OFF-HIGHWAY
 HIGHWAY VEHICLES
 SOLVENT UTILIZATION
 MISCELLANEOUS
 OTHER INDUSTRIAL PROCESSES
 FUEL COMB. INDUSTRIAL
 FUEL COMB. OTHER
 STORAGE & TRANSPORT
 FUEL COMB. ELEC. UTIL.
 CHEMICAL & ALLIED PRODUCT MFG
 PETROLEUM & RELATED INDUSTRIES



Test result: Passed

The following table illustrates test for county with maximum emission in year 2011 –

Table 13. Test for county with maximum emission in year 2011

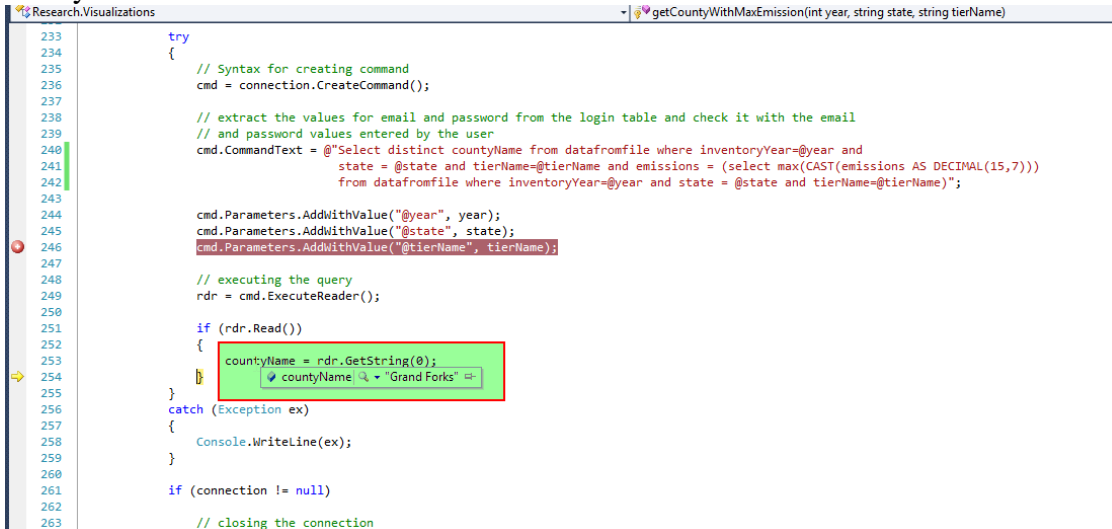
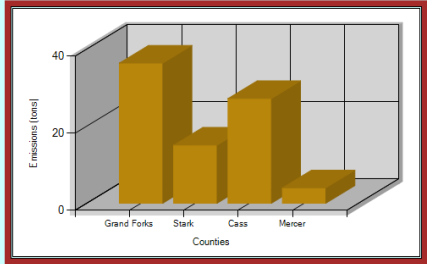
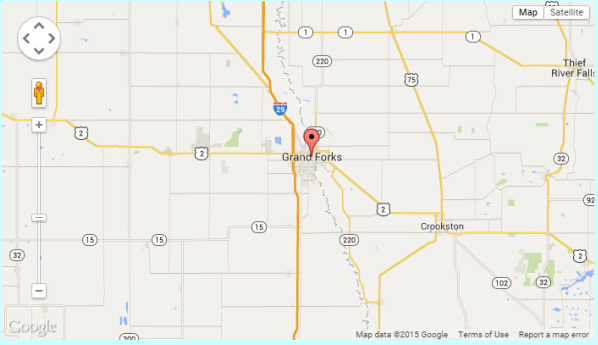
Display Name – More Visualizations for Emissions data	
Test 11	To verify if the test passes when for the year 2011, the county having the maximum emission has the highest bar in the bar chart and that the pin is being dropped at the approximate county location on the map and the expected result and the actual result is the same county location on the map having the highest bar in the bar chart
	<p>Input: Year: 2011 State: ND Emission Category Type: CHEMICAL & ALLIED PRODUCT MFG Output:</p> <p>county: Grand Forks</p> <p>Expected Result: county: Grand Forks</p>  <pre> 233 try 234 { 235 // Syntax for creating command 236 cmd = connection.CreateCommand(); 237 238 // extract the values for email and password from the login table and check it with the email 239 // and password values entered by the user 240 cmd.CommandText = @"Select distinct countyName from datafromfile where inventoryYear=@year and 241 state = @state and tierName=@tierName and emissions = (select max(CAST(emissions AS DECIMAL(15,7))) 242 from datafromfile where inventoryYear=@year and state = @state and tierName=@tierName)"; 243 244 cmd.Parameters.AddWithValue("@year", year); 245 cmd.Parameters.AddWithValue("@state", state); 246 cmd.Parameters.AddWithValue("@tierName", tierName); 247 248 // executing the query 249 rdr = cmd.ExecuteReader(); 250 251 if (rdr.Read()) 252 { 253 countyName = rdr.GetString(0); 254 consoleName = "Grand Forks"; 255 } 256 } catch (Exception ex) 257 { 258 Console.WriteLine(ex); 259 } 260 261 if (connection != null) 262 263 // closing the connection </pre>

Table 13. Test for county with maximum emission in year 2011 (continued)

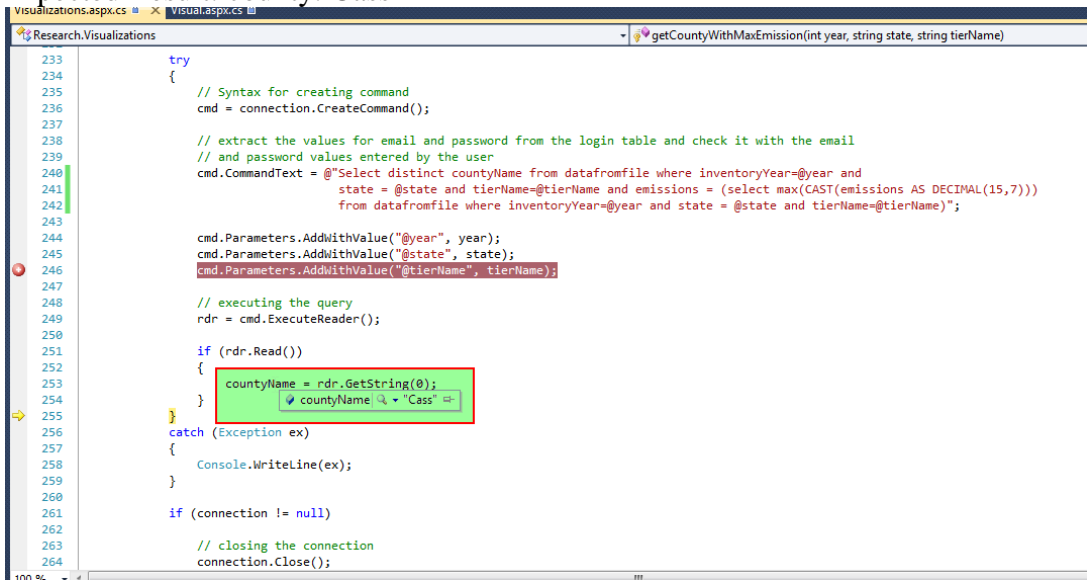
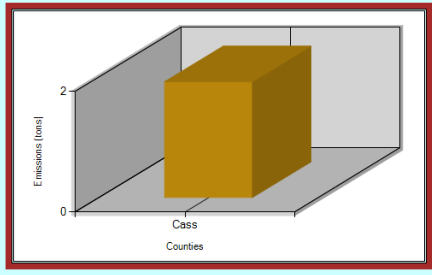

	<p>Actual Result: county: Grand Forks</p> <div style="background-color: #e0f7fa; padding: 10px; border: 1px solid black;"> <p style="text-align: center; color: #0070c0;"><i>More Visualizations for Emissions data</i></p> <p style="text-align: center; color: #0070c0;"><i>This data is for year: 2011</i></p> <p>Please select a State: <input type="text" value="ND"/> Home Visual Correlation</p> <p>Please select a Category Tier: <input type="text" value="CHEMICAL & ALLIED PRODUCT MFG"/> Home Visual Correlation</p> <p style="text-align: center;"><input type="button" value="Generate Visualization"/></p> <div style="display: flex; justify-content: space-around;"> <div style="border: 2px solid red; padding: 5px;">  <table border="1" style="display: none;"> <caption>3D Bar Chart Data (Approximate)</caption> <thead> <tr> <th>County</th> <th>Emissions (tons)</th> </tr> </thead> <tbody> <tr> <td>Grand Forks</td> <td>40</td> </tr> <tr> <td>Stark</td> <td>20</td> </tr> <tr> <td>Cass</td> <td>30</td> </tr> <tr> <td>Mercer</td> <td>10</td> </tr> </tbody> </table> </div> <div>  </div> </div> </div>	County	Emissions (tons)	Grand Forks	40	Stark	20	Cass	30	Mercer	10
County	Emissions (tons)										
Grand Forks	40										
Stark	20										
Cass	30										
Mercer	10										
	<p>Test result: Passed</p>										

The following table illustrates test for county with maximum emission in year 2008 –

Table 14. Test for county with maximum emission in year 2008

Display Name – More Visualizations for Emissions data	
Test 12	To verify if the test passes when for the year 2008, the county having the maximum emission has the highest bar in the bar chart and that the pin is being dropped at the approximate county location on the map and the expected result and the actual result is the same county location on the map having the highest bar in the bar chart
	<p>Input: Year: 2008 State: ND Emission Category Type: SOLVENT UTILIZATION Output: county: Cass</p>

Table 14. Test for county with maximum emission in year 2008 (continued)

	<p>Expected Result: county: Cass</p>  <pre> 233 try 234 { 235 // Syntax for creating command 236 cmd = connection.CreateCommand(); 237 238 // extract the values for email and password from the login table and check it with the email 239 // and password values entered by the user 240 cmd.CommandText = @"Select distinct countyName from datafromfile where inventoryYear=@year and 241 state = @state and tierName=@tierName and emissions = (select max(CAST(emissions AS DECIMAL(15,7))) 242 from datafromfile where inventoryYear=@year and state = @state and tierName=@tierName)"; 243 244 cmd.Parameters.AddWithValue("@year", year); 245 cmd.Parameters.AddWithValue("@state", state); 246 cmd.Parameters.AddWithValue("@tierName", tierName); 247 248 // executing the query 249 rdr = cmd.ExecuteReader(); 250 251 if (rdr.Read()) 252 { 253 countyName = rdr.GetString(0); 254 } 255 } 256 catch (Exception ex) 257 { 258 Console.WriteLine(ex); 259 } 260 261 if (connection != null) 262 { 263 // closing the connection 264 connection.Close(); </pre>
	<p>Actual Result: county: Cass</p>
	<p style="text-align: center;"><i>More Visualizations for Emissions data</i></p> <p style="text-align: center;"><i>This data is for year: 2008</i></p> <p>Please select a State: <input type="text" value="ND"/> Home Visual Correlation</p> <p>Please select a Category Tier: <input type="text" value="SOLVENT UTILIZATION"/></p> <p style="text-align: center;"><input type="button" value="Generate Visualization"/></p> <div style="display: flex; justify-content: space-around;"> <div data-bbox="332 1228 760 1501" style="border: 2px solid red; padding: 5px;">  </div> <div data-bbox="795 1186 1396 1543">  </div> </div>
	<p>Test result: Passed</p>

The following table illustrates test correlation coefficient for same states and same year –

Table 15. Test correlation coefficient for same states and same year

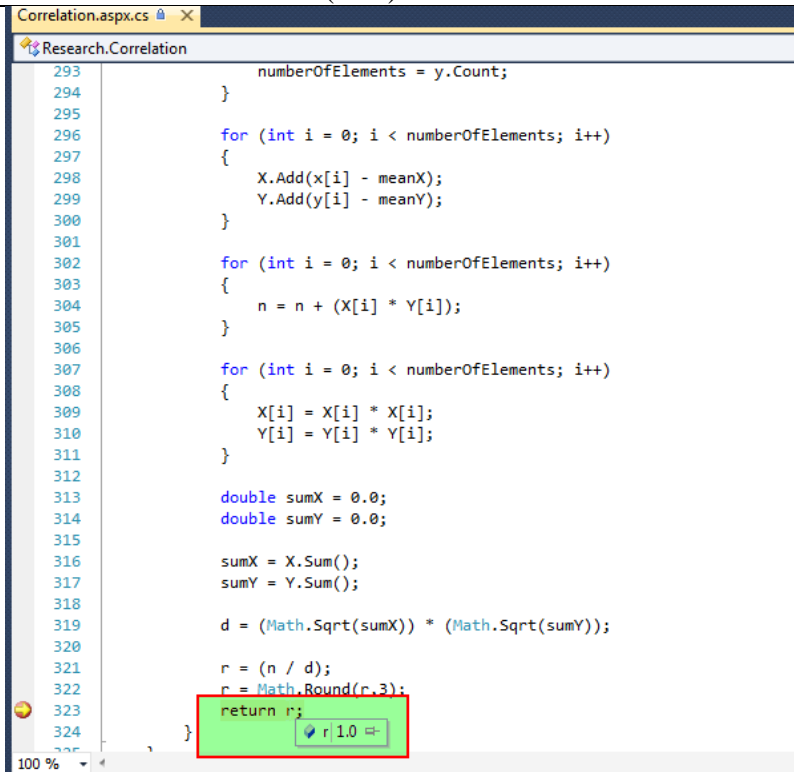
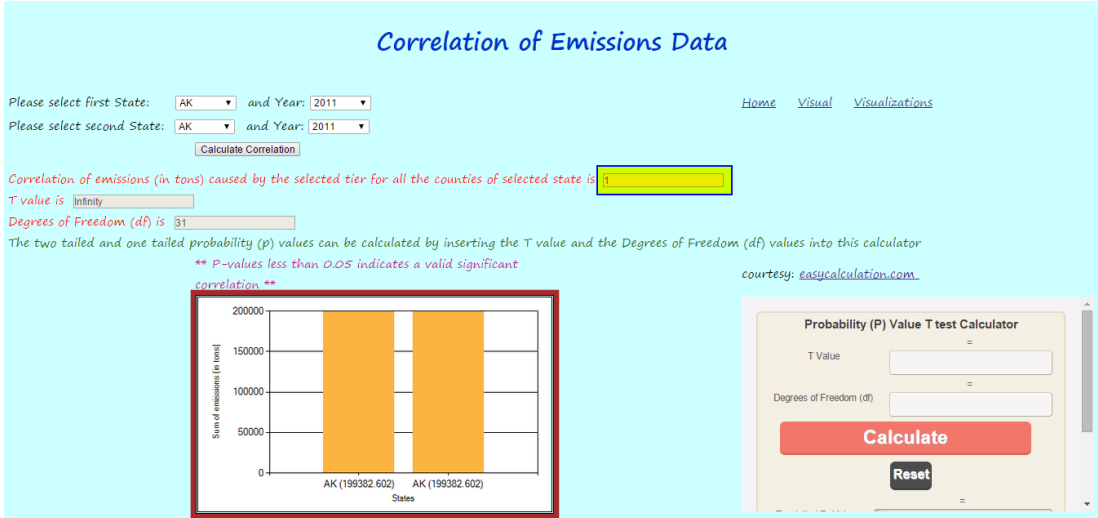
Display Name – Correlation of Emissions data	
Test 13	To verify if the test passes when the values for both the states are same and the values for both the years are same and the expected result and the actual result is 1 and that both the states are highly correlated
	Input: State1: AK Year1: 2011 State2: AK Year2: 2011 Output: 1
	Expected Result: Correlation of emissions (tons): 1
	 <pre> 293 numberOfElements = y.Count; 294 } 295 296 for (int i = 0; i < numberOfElements; i++) 297 { 298 X.Add(x[i] - meanX); 299 Y.Add(y[i] - meanY); 300 } 301 302 for (int i = 0; i < numberOfElements; i++) 303 { 304 n = n + (X[i] * Y[i]); 305 } 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r, 3); 323 return r; 324 } </pre>

Table 15. Test correlation coefficient for same states and same year (continued)

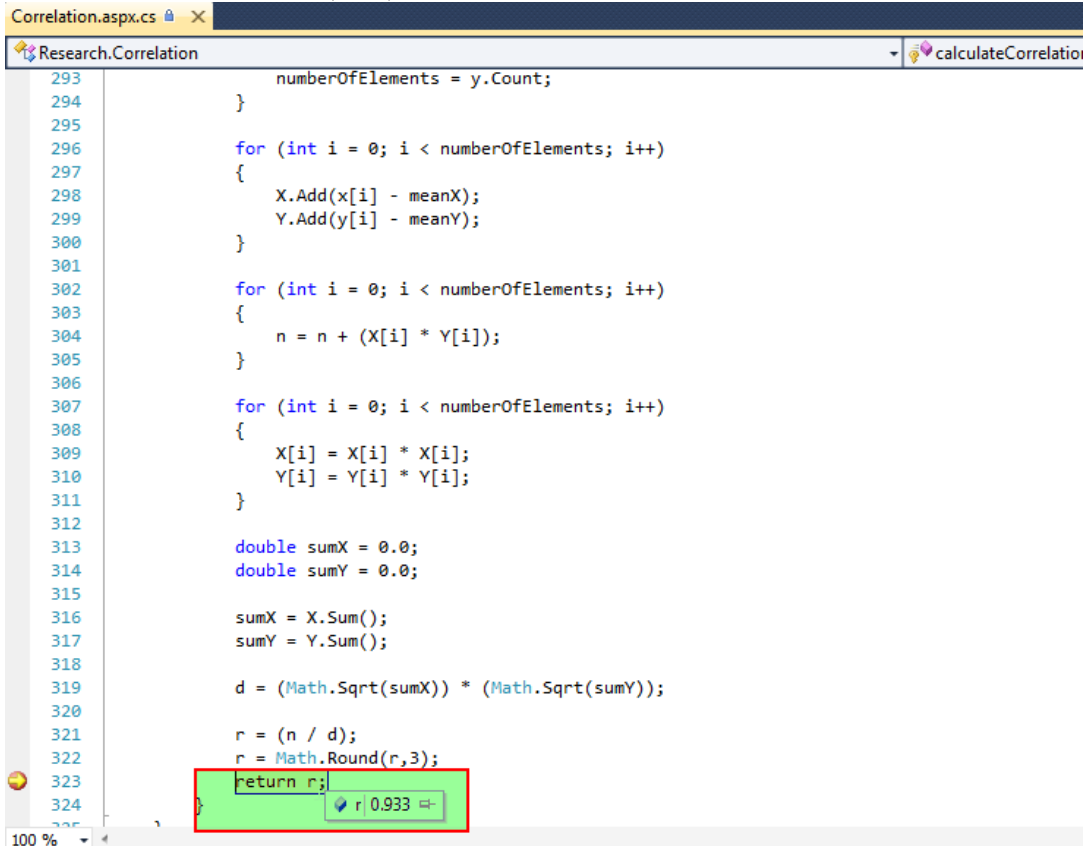
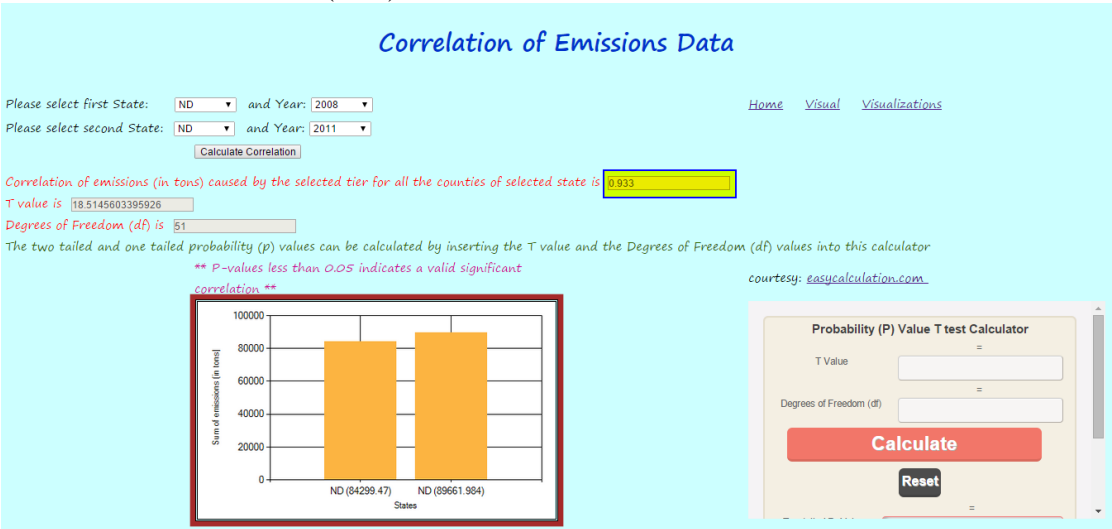
	<p>Actual Result: Correlation of emissions (tons): 1</p>  <p>Test result: Passed</p>
--	--

The following table illustrates test correlation coefficient for same states and different years –

Table 16. Test correlation coefficient for same states and different years

Display Name – Correlation of Emissions data	
Test 14	To verify if the test passes when the values for both the states are same and the values for both the years are different and the expected result and the actual result is a positive value and that the emissions for both the states are correlated
	<p>Input: State1: ND Year1: 2008 State2: ND Year2: 2011 Output: 0.933</p>

Table 16. Test correlation coefficient for same states and different years (continued)

	<p>Expected Result: Correlation of emissions (tons): 0.933</p> 
	<p>Actual Result: Correlation of emissions (tons): 0.933</p>  <p>Test result: Passed</p>

The following table illustrates test for negative correlation for same states and different years –

Table 17. Test for negative correlation for same states and different years

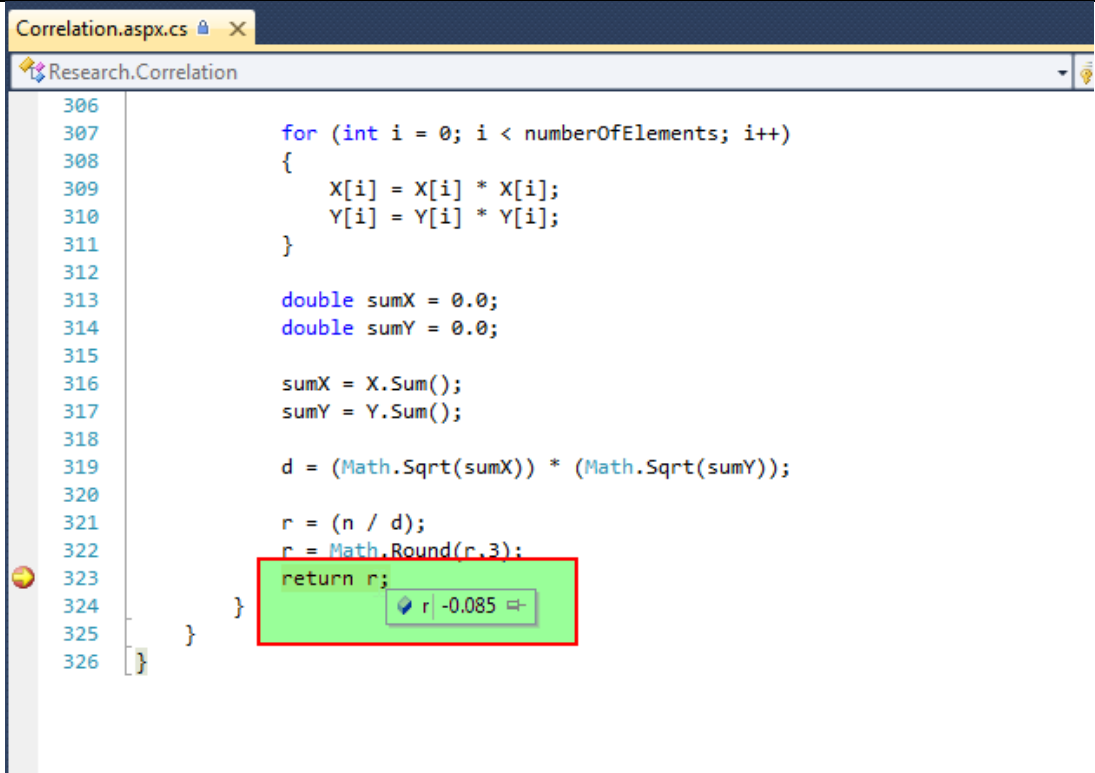
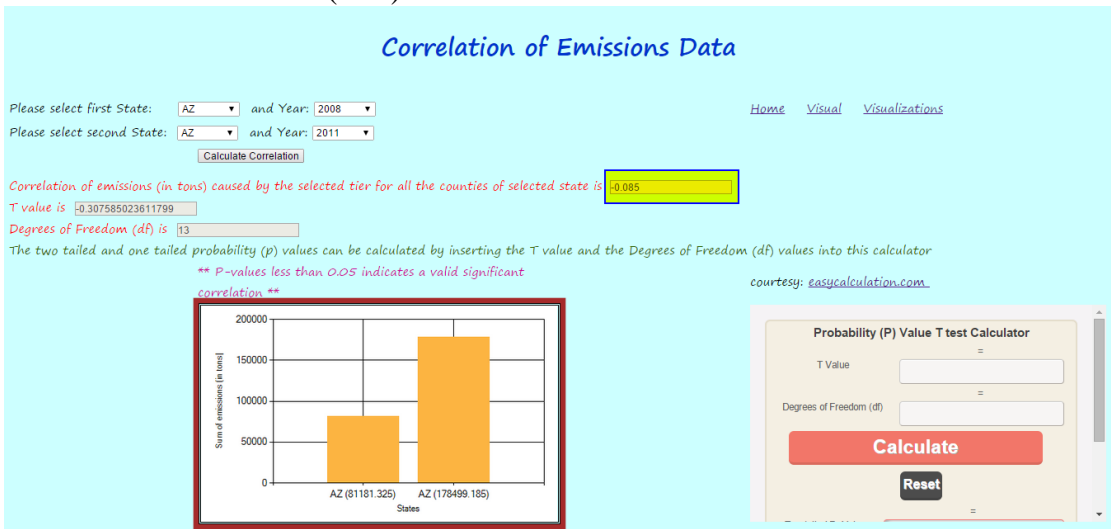
Display Name – Correlation of Emissions data	
Test 15	To verify if the test passes when the values for both the states are same and the values for both the years are different and the expected result and the actual result is a negative value and that the emissions from both the states cannot be correlated
	<p>Input: State1: AZ Year1: 2008 State2: AZ Year2: 2011 Output: -0.085</p>
	<p>Expected Result: Correlation of emissions (tons): -0.085</p>
	 <pre> 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r, 3); 323 return r; 324 } 325 326 </pre>

Table 17. Test for negative correlation for same states and different years (continued)

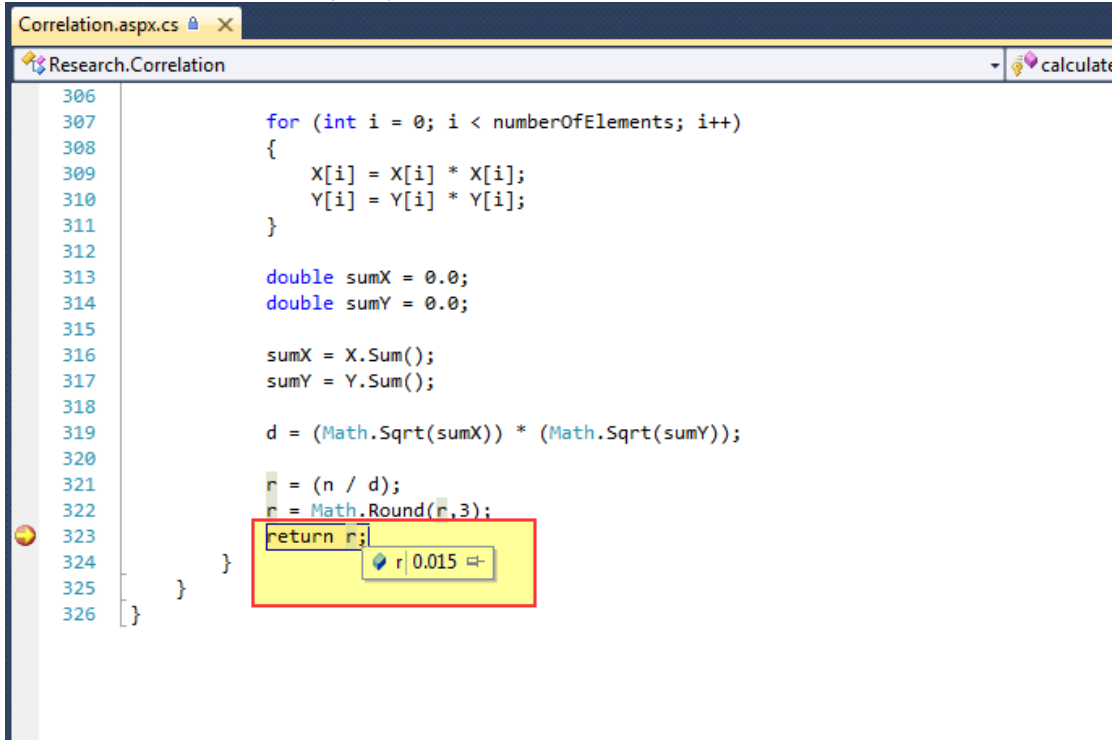
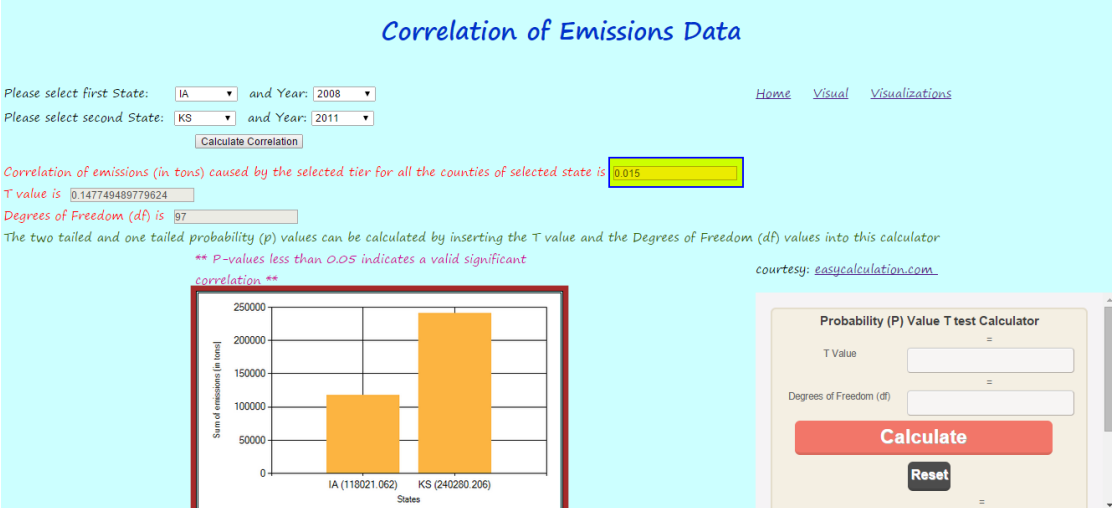
	<p>Actual Result: Correlation of emissions (tons): -0.085</p>  <p>Test result: Passed</p>
--	---

The following table illustrates test for positive correlation for different states and different years –

Table 18. Test for positive correlation for different states and different years

Display Name – Correlation of Emissions data	
Test 16	To verify if the test passes when the values for both the states are different and the values for both the years are different and the expected result and the actual result is a positive value and that the emissions for both the states are correlated
	<p>Input: State1: IA Year1: 2008 State2: KS Year2: 2011 Output: 0.015</p>

Table 18. Test for positive correlation for different states and different years (continued)

	<p>Expected Result: Correlation of emissions (tons): 0.015</p>  <pre> 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r, 3); 323 return r; 324 } 325 } 326 </pre>						
	<p>Actual Result: Correlation of emissions (tons): 0.015</p>						
	 <p>Correlation of Emissions Data</p> <p>Please select first State: IA and Year: 2008</p> <p>Please select second State: KS and Year: 2011</p> <p>Calculate Correlation</p> <p>Correlation of emissions (in tons) caused by the selected tier for all the counties of selected state is: 0.015</p> <p>T value is: 0.147749489779624</p> <p>Degrees of Freedom (df) is: 97</p> <p>The two tailed and one tailed probability (p) values can be calculated by inserting the T value and the Degrees of Freedom (df) values into this calculator</p> <p>** P-values less than 0.05 indicates a valid significant correlation **</p> <p>courtesy: easycalculation.com</p> <p>Sum of emissions (in tons)</p> <table border="1"> <thead> <tr> <th>State</th> <th>Sum of emissions (in tons)</th> </tr> </thead> <tbody> <tr> <td>IA (118021.062)</td> <td>118021.062</td> </tr> <tr> <td>KS (240280.206)</td> <td>240280.206</td> </tr> </tbody> </table> <p>Probability (P) Value T test Calculator</p> <p>T Value: <input type="text"/></p> <p>Degrees of Freedom (df): <input type="text"/></p> <p>Calculate</p> <p>Reset</p>	State	Sum of emissions (in tons)	IA (118021.062)	118021.062	KS (240280.206)	240280.206
State	Sum of emissions (in tons)						
IA (118021.062)	118021.062						
KS (240280.206)	240280.206						
	<p>Test result: Passed</p>						

The following table illustrates test for negative correlation for different states and different years –

Table 19. Test for negative correlation for different states and different years

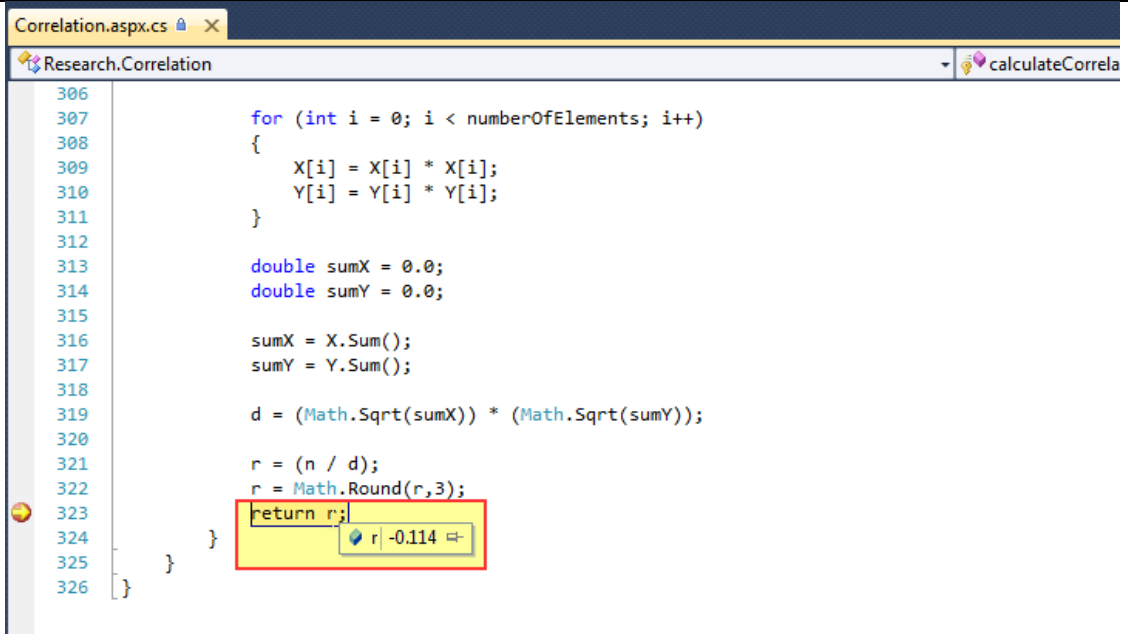
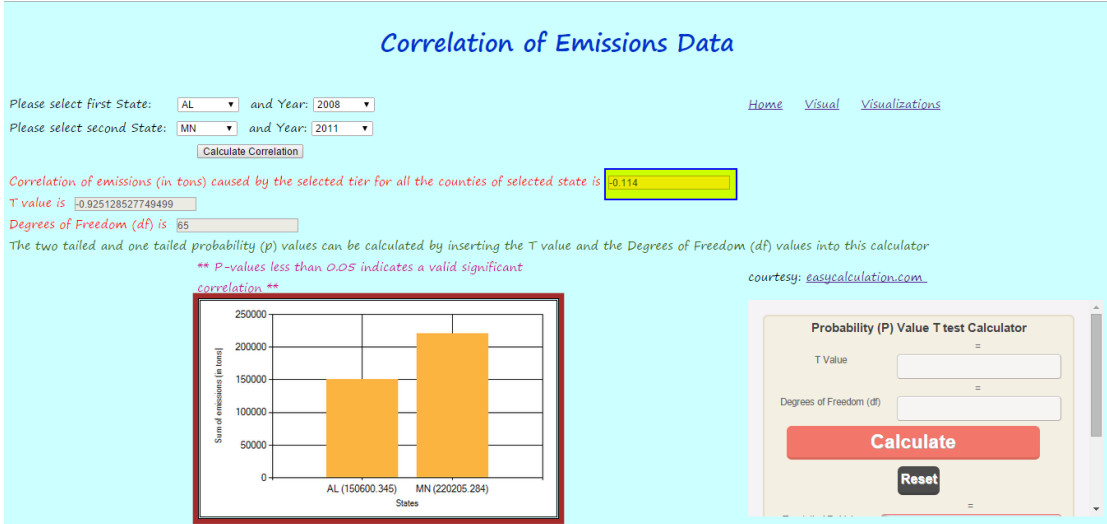
Display Name – Correlation of Emissions data	
Test 17	To verify if the test passes when the values for both the states are different and the values for both the years are different and the expected result and the actual result is a negative value and that both the states cannot be correlated
	<p>Input: State1: AL Year1: 2008 State2: MN Year2: 2011 Output: -0.114</p> <p>Expected Result: Correlation of emissions (tons): -0.114</p>
	 <pre> 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r,3); 323 return r; 324 } 325 326 </pre>

Table 19. Test for negative correlation for different states and different years (continued)

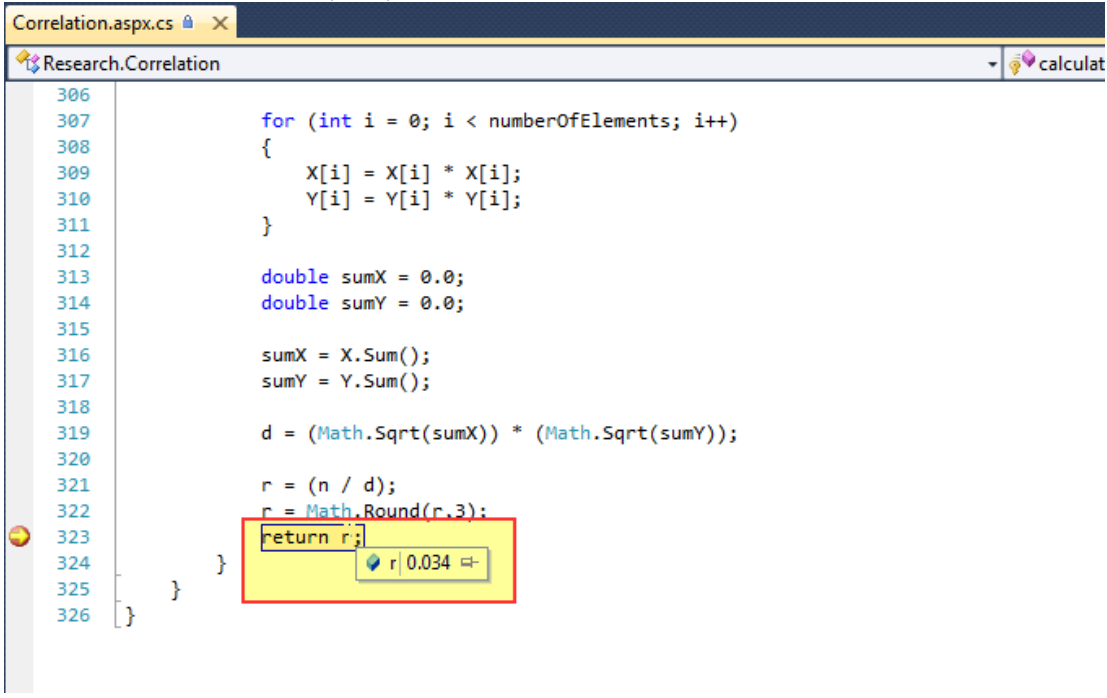
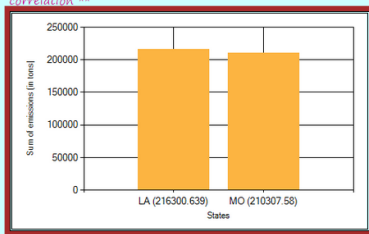
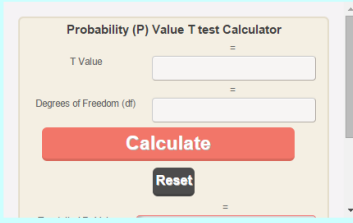
	<p>Actual Result: Correlation of emissions (tons): -0.114</p>  <p>Test result: Passed</p>
--	---

The following table illustrates test for positive correlation for different states and same year –

Table 20. Test for positive correlation for different states and same year

Display Name – Correlation of Emissions data	
Test 18	To verify if the test passes when the values for both the states are different and the values for both the years are same and the expected result and the actual result is a negative value and that the emissions for both the states are correlated
	Input: State1: LA Year1: 2011 State2: MO Year2: 2011 Output: 0.034

Table 20. Test for positive correlation for different states and same year (continued)

	<p>Expected Result: Correlation of emissions (tons): 0.034</p>  <pre> 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r,3); 323 return r; 324 } 325 326 </pre>
	<p>Actual Result: Correlation of emissions (tons): 0.034</p>
	<p style="text-align: center;"><i>Correlation of Emissions Data</i></p> <p>Please select first State: <input type="text" value="LA"/> and Year: <input type="text" value="2011"/></p> <p>Please select second State: <input type="text" value="MO"/> and Year: <input type="text" value="2011"/></p> <p style="text-align: center;"><input type="button" value="Calculate Correlation"/></p> <p>Correlation of emissions (in tons) caused by the selected tier for all the counties of selected state is <input type="text" value="0.034"/></p> <p>T value is <input type="text" value="0.27002275201048"/></p> <p>Degrees of Freedom (df) is <input type="text" value="63"/></p> <p>The two tailed and one tailed probability (p) values can be calculated by inserting the T value and the Degrees of Freedom (df) values into this calculator ** P-values less than 0.05 indicates a valid significant correlation ** courtesy: easycalculation.com</p>   <p>Test result: Passed</p>

The following table illustrates test for negative correlation for different states and same year –

Table 21. Test for negative correlation for different states and same year

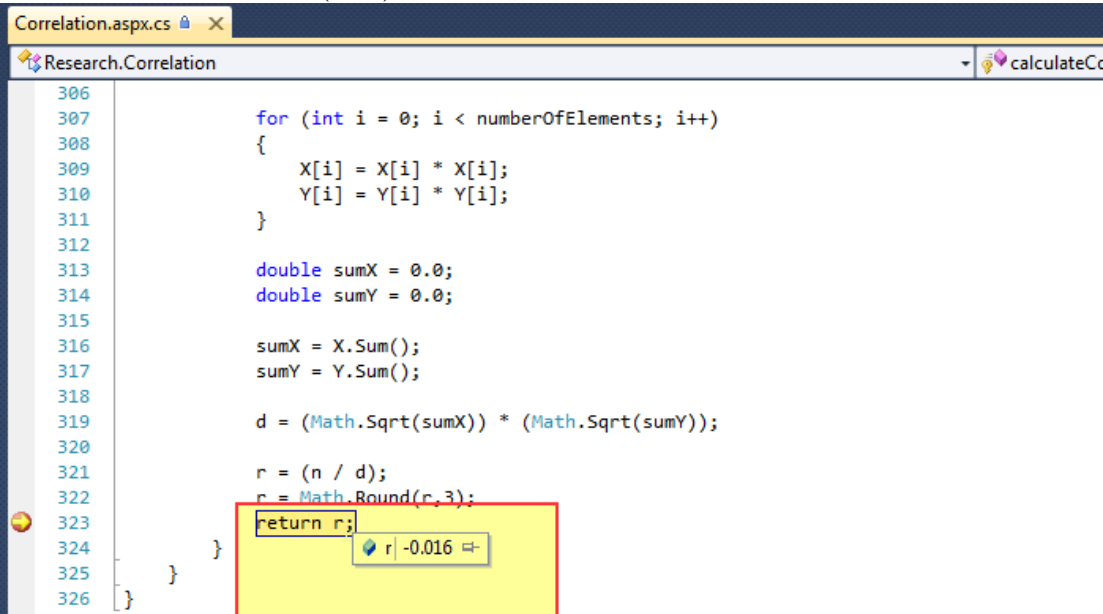
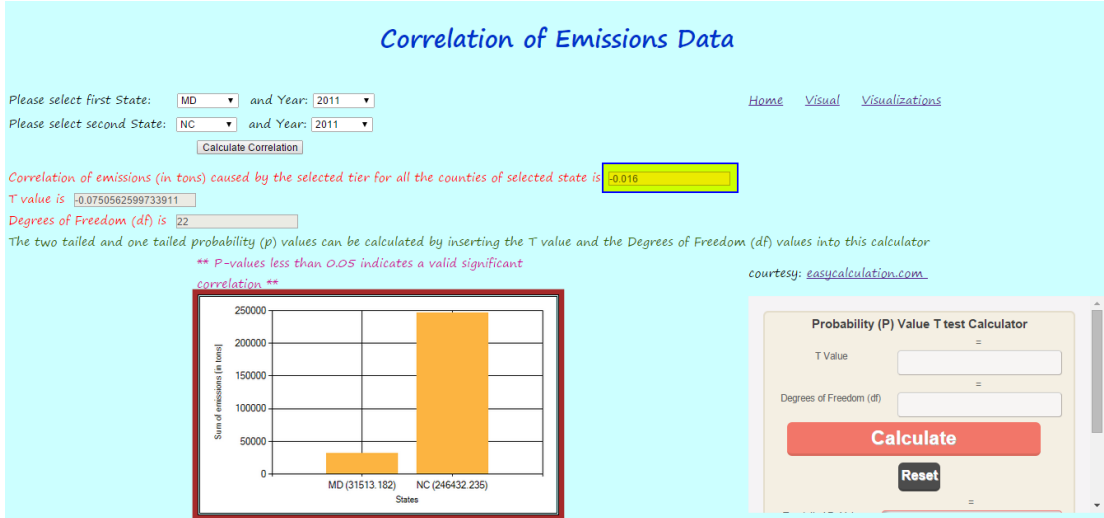
Display Name – Correlation of Emissions data	
Test 19	To verify if the test passes when the values for both the states are different and the values for both the years are same and the expected result and the actual result is a negative value and that both the states cannot be correlated
	<p>Input: State1: MD Year1: 2011 State2: NC Year2: 2011 Output: -0.016</p> <p>Expected Result: Correlation of emissions (tons): -0.016</p>  <pre> 306 307 for (int i = 0; i < numberOfElements; i++) 308 { 309 X[i] = X[i] * X[i]; 310 Y[i] = Y[i] * Y[i]; 311 } 312 313 double sumX = 0.0; 314 double sumY = 0.0; 315 316 sumX = X.Sum(); 317 sumY = Y.Sum(); 318 319 d = (Math.Sqrt(sumX)) * (Math.Sqrt(sumY)); 320 321 r = (n / d); 322 r = Math.Round(r, 3); 323 return r; 324 } 325 } 326 } </pre>

Table 21. Test for negative correlation for different states and same year (continued)

	<p>Actual Result: Correlation of emissions (tons): -0.016</p>  <p>Test result: Passed</p>
--	---

The following table illustrates test for positive correlation for different states and same year –

Table 22. Test for positive correlation for different states and same year

Display Name – Correlation of Emissions data	
Test 20	To verify if the test passes when the total sum of emissions values for both the states for given years being displayed on the bar chart match the total sum of emissions values in the database and the expected result and the actual result are the same states with the same values as per the database
	<p>Input: State1: ND Year1: 2011 State2: KS Year2: 2011</p> <p>Output: State1: ND Sum of Emissions (tons): 89661.984 State2: KS Sum of Emissions (tons): 240280.206</p>

Table 22. Test for positive correlation for different states and same year (continued)

Expected Result:

State1: ND

Sum of Emissions (tons): 89661.984

State2: KS

Sum of Emissions (tons): 240280.206

```

256         {
257             emissions = rdr.GetString(0);
258         }
259     }
260     catch (Exception ex)
261     {
262         Console.WriteLine(ex);
263     }
264
265     if (connection != null)
266     {
267         // closing the connection
268         connection.Close();
269     }
270     return emissions;
271 }

```

emissions | 89661.984

// this method calculates correlation when two lists of all the total emissions in individual counties caused by
// emissions from all tier types are provided.
protected double calculateCorrelation(List<double> x, List<double> y)
{
double r = 0.0, n = 0.0, d = 0.0;
double meanX = 0.0;
double meanY = 0.0;
List<double> x = new List<double>();

```

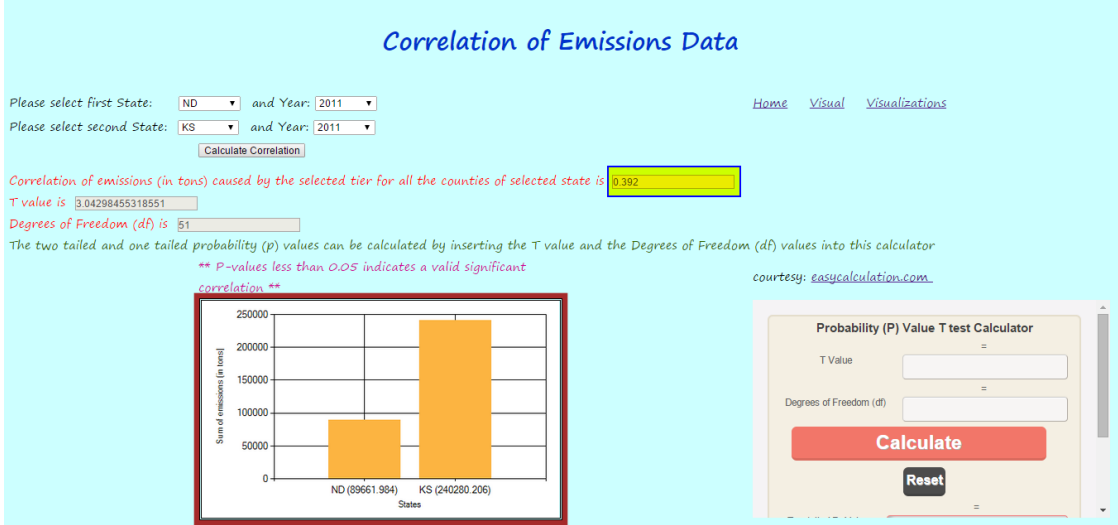
246 // datafromfile table and check it with the year and state values entered by the user
247 cmd.CommandText = "select ROUND(sum(emissions),3) from datafromfile where state=@state and inventor
248
249 cmd.Parameters.AddWithValue("@state", state);
250 cmd.Parameters.AddWithValue("@year", year);
251
252 // executing the query
253 rdr = cmd.ExecuteReader();
254
255 if (rdr.Read())
256 {
257     emissions = rdr.GetString(0);
258 }
259 }
260 catch (Exception ex)
261 {
262     Console.WriteLine(ex);
263 }
264
265 if (connection != null)
266 {
267     // closing the connection
268     connection.Close();
269 }
270 return emissions;
271 }

```

emissions | 240280.206

// this method calculates correlation when two lists of all the total emissions in individual counties caus
// emissions from all tier types are provided.
protected double calculateCorrelation(List<double> x, List<double> y)
{
double r = 0.0, n = 0.0, d = 0.0;

Table 22. Test for positive correlation for different states and same year (continued)

	<p>Actual Result: State1: ND Sum of Emissions (tons): 89661.984 State2: KS Sum of Emissions (tons): 240280.206</p>  <p>Test result: Passed</p>
--	--

The following table illustrates test for number of days with pm2.5 and number of days with AQI for a state –

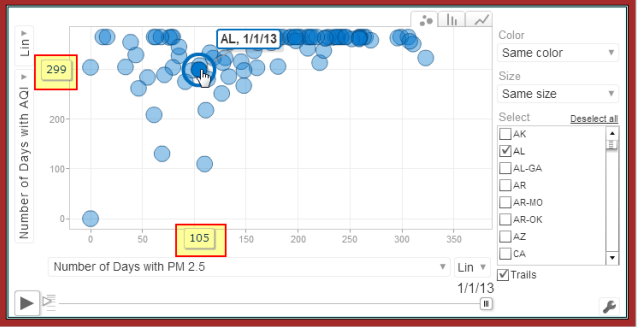
Table 23. Test for number of days with pm2.5 and number of days with AQI for a state

Display Name – Motion Chart	
Test 21	To verify if the test passes when the number of days with pm2.5 and number of days with AQI values for a particular state for a particular time matches the value in the database and the expected result and the actual result is the same state with the same values as per the database
	Input: State: AL Year: 2013 Output: Number of days with pm2.5: 105 Number of days with AQI: 299

Table 23. Test for number of days with pm2.5 and number of days with AQI for a state (continued)

Expected Result:
 Number of days with pm2.5: 105
 Number of days with AQI: 299

Table 23. Test for number of days with pm2.5 and number of days with AQI for a state (continued)

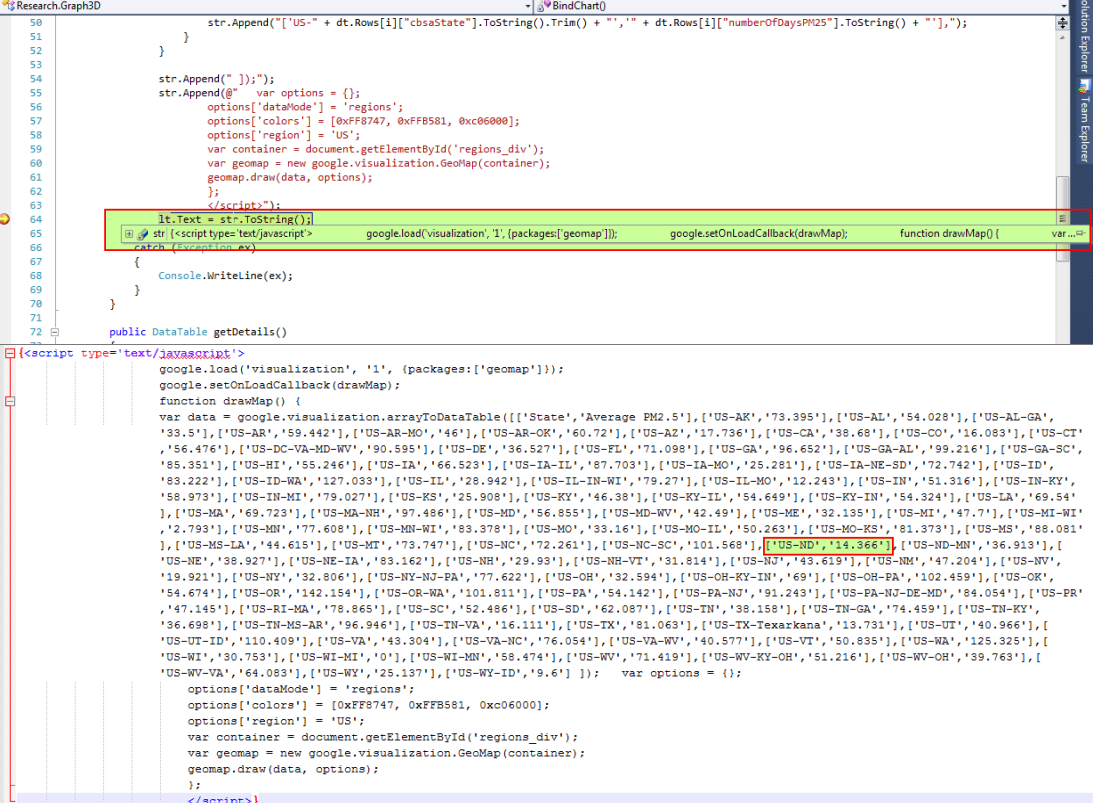
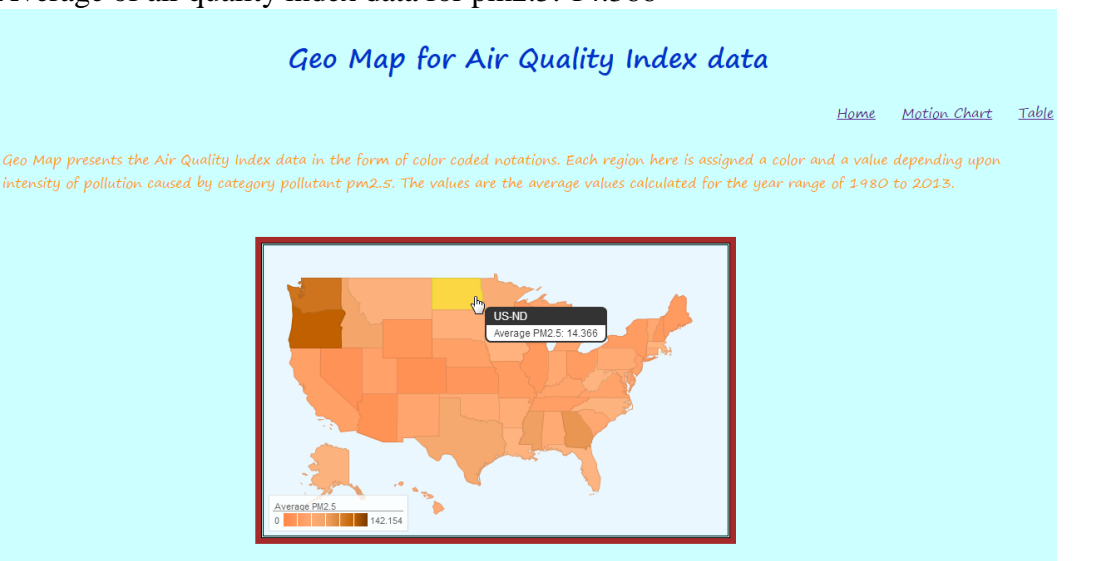
	<p>Actual Result: Number of days with pm2.5: 105 Number of days with AQI: 299</p> <div style="background-color: #e0f7fa; padding: 10px; border: 1px solid #cfcfcf;"> <p style="text-align: center; color: #0070c0; font-weight: bold;">Motion Chart for Air Quality Index data</p> <p style="text-align: right;">Home Geo Map Table</p> <p style="color: #a020f0; font-size: small;">Motion chart will let you analyze and visualize Air Quality Index data dynamically over time range of 1980 to 2013. Increasing intensity of pollution caused by category pm2.5 can be observed here.</p>  </div> <p>Test result: Passed</p>
--	--

The following table illustrates test for average AQI data for pm2.5 for highlighted state –

Table 24. Test for average AQI data for pm2.5 for highlighted state

Display Name – Geo Map	
Test 22	To verify if the test passes when the average of air quality index data for pm2.5 for the highlighted state matches the value in the database and the expected result and the actual result matches the calculated average value as per the database
	Input: State: ND Output: State: ND Average of air quality index data for pm2.5: 14.366

Table 24. Test for average AQI data for pm2.5 for highlighted state (continued)

	<p>Expected Result: State: ND Average of air quality index data for pm2.5: 14.366</p>
	 <pre> google.load('visualization', '1', {packages:['geomap']}); google.setOnLoadCallback(drawMap); function drawMap() { var data = google.visualization.arrayToDataTable([['State','Average PM2.5'], ['US-AK','73.395'], ['US-AL','54.028'], ['US-AL-GA', '33.5'], ['US-AR','59.442'], ['US-AR-MO','46'], ['US-AR-OK','60.72'], ['US-AZ','17.736'], ['US-CA','38.68'], ['US-CO','16.083'], ['US-CT', '56.476'], ['US-DC-VA-MD-WV','90.595'], ['US-DE','36.527'], ['US-FL','71.098'], ['US-GA','96.652'], ['US-GA-AL','99.216'], ['US-GA-SC', '85.351'], ['US-HI','55.246'], ['US-IA','66.523'], ['US-IA-IL','87.703'], ['US-IA-MO','25.281'], ['US-IA-NE-SD','72.742'], ['US-ID', '83.222'], ['US-ID-WA','127.033'], ['US-IL','28.942'], ['US-IL-IN-WI','79.27'], ['US-IL-MO','12.243'], ['US-IN','51.316'], ['US-IN-KY', '58.973'], ['US-IN-MI','79.027'], ['US-KS','25.908'], ['US-KY','46.38'], ['US-KY-IL','54.649'], ['US-KY-IN','54.324'], ['US-LA','69.54', '1'], ['US-MA','69.723'], ['US-MA-NH','97.486'], ['US-MD','56.855'], ['US-MD-WV','42.49'], ['US-ME','32.135'], ['US-MI','47.7'], ['US-MI-WI', '2.793'], ['US-MN','77.608'], ['US-MN-WI','83.378'], ['US-MO','33.16'], ['US-MO-IL','50.263'], ['US-MO-KS','81.373'], ['US-MS','88.081', '1'], ['US-MS-LA','44.615'], ['US-MT','73.747'], ['US-NC','72.261'], ['US-NC-SC','101.568'], ['US-ND','14.366'], ['US-ND-MN','36.913'], ['US-NE','38.927'], ['US-NE-IA','83.162'], ['US-NH','29.93'], ['US-NH-VT','31.814'], ['US-NJ','43.619'], ['US-NM','47.204'], ['US-NV', '19.921'], ['US-NY','32.806'], ['US-NY-NJ-PA','77.622'], ['US-OH','32.594'], ['US-OH-KY-IN','69'], ['US-OH-PA','102.459'], ['US-OK', '54.674'], ['US-OR','142.154'], ['US-OR-WA','101.811'], ['US-PA','54.142'], ['US-PA-NJ','91.243'], ['US-PA-NJ-DE-MD','84.054'], ['US-PR', '47.145'], ['US-RI-MA','78.865'], ['US-SC','52.486'], ['US-SD','62.087'], ['US-TN','38.158'], ['US-TN-GA','74.459'], ['US-TN-KY', '36.698'], ['US-TN-MS-AR','96.946'], ['US-TN-VA','116.111'], ['US-TX','81.063'], ['US-TX-Texarkana','13.731'], ['US-UT','40.966'], ['US-UT-ID','110.409'], ['US-VA','43.304'], ['US-VA-NC','76.054'], ['US-VA-WV','40.577'], ['US-VI','50.835'], ['US-WA','125.325'], ['US-WI','30.753'], ['US-WI-MI','0'], ['US-WI-MN','58.474'], ['US-WV','71.419'], ['US-WV-KY-OH','51.216'], ['US-WV-OH','39.763'], ['US-WV-VA','64.083'], ['US-WY','25.137'], ['US-WY-ID','9.6']]); var options = {}; options['dataMode'] = 'regions'; options['colors'] = [0xFF8747, 0xFFB581, 0xc06000]; options['region'] = 'US'; var container = document.getElementById('regions_div'); var geomap = new google.visualization.GeoMap(container); geomap.draw(data, options); } </pre>
	<p>Actual Result: State: ND Average of air quality index data for pm2.5: 14.366</p>  <p>Geo Map for Air Quality Index data</p> <p>Home Motion Chart Table</p> <p>Geo Map presents the Air Quality Index data in the form of color coded notations. Each region here is assigned a color and a value depending upon intensity of pollution caused by category pollutant pm2.5. The values are the average values calculated for the year range of 1980 to 2013.</p>
	<p>Test result: Passed</p>

CHAPTER 7. DISCUSSION

The completeness of PAS application can be judged by the amount of work done to analyze and visualize the data in terms of different visual representations and the usage of correlation techniques to further aid the understanding of these graphical and visual representations. These constitute the strengths of the systems which are further elaborated in this chapter. Along with the strengths there were also some limitations of this system which are also discussed in this chapter.

7.1. Strengths

Interactive visualizations are a powerful and efficient way to represent and visualize data. In situations where it is difficult to read, interpret and comprehend large amounts of data in spreadsheets and reports, visualizations can help to effectively give meaning to the data. The visualizations supported are the strength of my research work since they hold potential for improving data interpretation and comprehension.

The Pollution Analysis System (PAS) can be used by users. Users include but are not limited to regular people, government officials and other private sector officials can benefit from these visualizations. The visualizations can alert the people living in pollution ridden places with high levels of pm2.5, and people traveling through such an area can be cautious and can take preventive measures. For example, emissions data for category pollutant pm2.5 as a result of pollution source fuel combustions is very high in Cass County in North Dakota. This information can alert students and other people and encourage them to either use bikes or fuel efficient modes of transportation. This way they can greatly reduce the chances of being adversely affected by pollution in category pm2.5.

This system also allows the user to understand and compare the impact of pollutant category pm2.5 among different states and different years or different counties within the same state. By looking at the interactive visualizations in terms of bar charts, graphs and maps, the user can deduce how the pollution levels, especially in category pm2.5, has impacted the environment of a place. The summarization of air quality index data via motion chart can help the user understand the trend of pollution containing category pm2.5 for different geographical locations and its impact on the quality of air over a wide period of time. This can help a user infer how well the pollution needs to be controlled. We note that in some cases states have experienced gradually increasing pollution levels over a period of time. If the government officials, other government or private agencies, or other state mandated pollution control facilities are accessing this data, they are empowered to adopt some preventive measures; find ways to control and effectively reduce the pollution levels; and help improve the quality of the air.

Correlation is an important aspect of PAS. Correlation techniques used in PAS can help the user correlate and figure out a linear association between the pollution data in different states for different years. Correlation coefficient calculated using two data sets first in one state for a year and then in a second state for a year, provides information to reveal positive or negative correlation. Significance and validity of the correlation coefficient provides the user with the confidence that the correlation coefficient results and the data sets used are valid and are not just a coincidence. It is also important to not confuse correlation with causation.

7.2. Limitations

PAS was created with several functionalities in mind. Although the features for representing elaborate data visualizations were carefully designed and implemented, there remain some limitations of the system. Some of these limitations are as follows –

- Geo Map for Air Quality Index data – Geo Map uses a standard color scale with different shades of a color hue. The color scale on the Geo Map does not have a variety of colors for further differentiation of the average Air Quality Index values for different states. Also, the hover text for different states only provides the state abbreviations and the average Air Quality Index value. The hover text does not contain the values for all the cities or counties of a state.
- Table for Air Quality Index data – The color coded thresholds for the Good or acceptable Air Quality Index or a healthy environment and the Unhealthy or unacceptable Air Quality Index are predefined. The user of the application does not have a choice to set these values manually according to personal preferences.
- Visualizations of Emissions data display – The map is bound to the State and the Category tier type (sources of pollution) dropdown values and drops the pin on the location of the county which has the highest pollution containing category pollutant pm2.5. The map does not display the county boundaries.
- Correlation – PAS does determine a linear association of two data sets in the form of correlating the two states for their pollution data for two different years, and calculates the correlation coefficient on the basis of the selections of states and years. The significance, confidence and validity of this correlation coefficient are confirmed by the

calculations of the p-value. However, the Pearson correlation is the only statistical analysis PAS does.

- Motion Chart – PAS displays the trend of the Air Quality Index data for different states in the range of years from 1980 to 2013 and solely relies on the data provided by the Environmental Protection Agency (EPA) website. The total number of days for which the Air Quality Index data was recorded and reported is not consistent for all the years for all the states. Hence, the average value for the total number of days in a year for which the Air Quality Index data was recorded and the total number of days out of these where pm2.5 was the main pollutant is being used in the motion chart renderings. For example, the total number of days where Air Quality Index data was recorded for Florence, SC for year 2012 was 309, out of which the total number of days when pm2.5 was main pollutant was 94.

Although PAS application has some features and functionalities which can very well analyze and visualize the data in terms of visual and graphical representations, the system has some limitations as well. Some of these limitations include wide range of color notations for the Geo Map, availability of more air quality index data for the motion chart, usage of more statistical analysis other than Pearson correlation, giving the user an option to set the color codes for the threshold values according to their personal preferences and including county boundaries on the map.

CHAPTER 8. CONCLUSION AND FUTURE WORK

PAS is a dedicated application which has done a careful analysis of pollutant pm2.5 data and has visualized it in terms of a variety of visual and graphical representations. Usage of correlation techniques has further aided the utility of this application. This chapter concludes the application features and functionalities and the amount of research work done to make this application an important tool to enhance the understanding of pollutant pm2.5 data.

8.1. Conclusion

The pollution data in category pm2.5 from several countries was gathered; analyzed and meaningful visualizations were defined as a result of it. Subsequently, appropriate programming language, tools and data management choices were made and a design and prototype of this geographical information system were created. The graphical user interface PAS was prototyped to demonstrate the analyzed data in terms of charts and maps was designed using ASP.Net with C# as a programming language paired with MySQL 5.6 database. MySQL 5.6 database was interfaced using MySQL 6.0 CE Workbench. The data chosen for analysis and visualization was the 2008 and 2011 National Emissions Inventory data for category pm2.5 taken from the U.S Environmental Protection Agency website. This data was analyzed, compared and correlated on the basis of states and the emission (in tons) of the pollutants in different category tier types. This visualization was furthered to the level of emission (in tons) in counties belonging to a particular state as a result of a category tier type.

Apart from the National Emissions Inventory data, Air Quality Index data was also taken into account and was visualized in terms of motion chart which represented the trend of air quality index ranging from 1980 to 2013, for several US cities and states. Also, all of the data was demonstrated in a tabular form which can be managed and viewed by the user on a yearly

basis. Geo chart representations of the data were also provided to the user to aid ease of understanding of the geographical locations and the intensities to which they are struck by pollutants in category pm2.5.

8.2. Future Work

It is anticipated that this application can be further enhanced when more data sets are received by the database including, and not limited to, US states, cities and counties but other countries as well. Dr. Nygard's promising prospective research will bring a constant and consistent stream of data pertaining to pollution in China containing pollutants in category pm2.5 being uploaded to the server instantly. After this data is being successfully retained in the database, more data visualizations can be fabricated. Since the data will be constantly streamed to the server, the Admin user of the application won't require frequent uploading of data files every time for the users to be able to visualize it. Not only that, with the availability of varied data sets, more correlation techniques can be used and the user will be able to correlate data in a lot of different ways. Also, when data from different countries is available, it can be further analyzed to the level of states/provinces/districts and visualized. Varied sets of data including, for instance, US (its states, cities or counties), China (its provinces and cities), India (its states and cities), etc, can evidently improve the current state of the PAS.

Furthermore, this application can be internally enhanced in terms of data and color coding selections on the Table for Air Quality Index data display page. The user may be given an option to set the high, moderate, average and good pm2.5 pollutant levels in the atmosphere which will allow the user to modify red and green highlights on the data in the table.

CHAPTER 9. REFERENCES

- [1] *About Gapminder*. Retrieved March 8, 2015, from Gapminder: <http://www.gapminder.org/about-gapminder/>
- [2] *About Gapminder*. Retrieved March 8, 2015, from Gapminder: <http://www.gapminder.org/about-gapminder/history/>
- [3] *Air Data*. (2014, October 8). Retrieved November 8, 2014, from EPA United States Environmental Protection Agency: http://www.epa.gov/airdata/ad_about_reports.html#aqi
- [4] Battista, V. C. (2011). Motion Charts: Telling Stories with Statistics. *Section on Statistical Graphics - JSM 2011* , 4473-4483.
- [5] *Beijing Air Pollution: Real-time Air Quality Index (AQI)*. Retrieved March 8, 2015, from ACIQN: <http://aqicn.org/city/beijing/>
- [6] *Data Visualization*. Retrieved March 8, 2015, from SAS: http://www.sas.com/en_us/insights/big-data/data-visualization.html
- [7] *Emission Inventories*. Retrieved April 16, 2014, from U.S. Environmental Protection Agency: <http://www.epa.gov/ttn/chief/eiinformation.html>
- [8] Engel-Cox, J. A. (2005). Application of Satellite Remote-Sensing Data for Source Analysis of Fine Particulate Matter Transport Events. *Journal of the Air & Waste Management Association* , 1389-1397.
- [9] *Fine Particulate Matter*. Retrieved 02 22, 2015, from Ontario Ministry of the Environment and Climate Change: <http://www.airqualityontario.com/science/pollutants/particulates.php>
- [10] *Gapminder a fact-based world view*. Retrieved March 8, 2015, from Gapminder: <http://www.gapminder.org/>
- [11] Lippmann M, T. G. (1996). Sulfate concentrations as an indicator or ambient particulate matter air pollution for health risk evaluations. *Journal of Exposure Analysis and Environmental Epidemiology* , 123-146.
- [12] *Manipulating Data on Linux*. (2012, September 27). Retrieved March 8, 2015, from University of California, Irvine: <http://moo.nac.uci.edu/~hjm/ManipulatingDataOnLinux.html>
- [13] Ostro, B., & Chestnut, L. (1998). Assessing the Health Benefits of Reducing Particulate Matter Air Pollution in the United States. *Environmental Research* , 94-106.
- [14] Peng, R. D. (2008). A Method for Visualizing Multivariate Time Series Data. *Journal of Statistical Software* .
- [15] Raffuse, S. M. (2005). Estimating Regional Contributions to Atmospheric Haze Using GIS. *Esri International User Conference Proceedings*. San Diego.

- [16] Snyder, E. G. (2013). The Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology* , 11369-11377.
- [17] Sun, Y., Zhuang, G., Tang, A., Wang, Y., & An, Z. (2006). Chemical Characteristics of PM_{2.5} and PM₁₀ in Haze-Fog Episodes in Beijing. *Environmental Science & Technology* , 3148-3155.
- [18] *The 2011 National Emissions Inventory*. (n.d.). Retrieved April 16, 2014, from U.S. Environmental Protection Agency: <http://www.epa.gov/ttn/chief/net/2011inventory.html>
- [19] *Trendalyzer*. (n.d.). Retrieved March 8, 2015, from Wikipedia: <http://en.wikipedia.org/wiki/Trendalyzer>
- [20] Stangroom, J. *Pearson Correlation Coefficient Calculator*. Retrieved April 8, 2015, from Social Science Statistics: <http://www.socscistatistics.com/tests/pearson/>
- [21] Stangroom, J. *Pearson Correlation Coefficient Calculator*. Retrieved April 8, 2015, from Social Science Statistics: <http://www.socscistatistics.com/tests/pearson/>
- [22] *Correlation (Pearson, Kendall Spearman)*. Retrieved April 8, 2015, from Statistics Solutions Advancement Through Clarity: <http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/>
- [23] *Pearson's Correlation*. Retrieved April 8, 2015, from <http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf>
- [24] Siegle, D. (2009, October 14). *Is there a relationship (difference) or insn't there a relationship (difference)?* Retrieved April 8, 2015, from <http://www.gifted.uconn.edu/siegle/research/correlation/alphaleve.htm>
- [25] Lowry, R. Appendix to Chapter 4 The Significance of a Correlation Coefficient. In R. Lowry, *Concepts and Applications of Inferential Statistics*.
- [26] *Probability (P) Value T test Calculator*. Retrieved April 8, 2015, from [easycalculation.com: https://www.easycalculation.com/statistics/p-value-t-test.php](https://www.easycalculation.com/statistics/p-value-t-test.php)
- [27] *Interpreting the Pearson coefficient*. Retrieved April 8, 2015, from OPEX Resources: <http://www.opexresources.com/index.php/free-resources/articles/interpreting-the-pearson-coefficient>

APPENDIX

In Software Engineering, use cases are described as a list of steps which define the interaction between a role and the system in order to achieve a particular goal. In terms of Unified Modular Language (UML), a role is known as an actor which can be a human or time or an external system. Through the series of these use cases, an attempt to describe the function and working of Pollution Analysis System (PAS) has been made.

Table A1. Use case to describe the Admin user Home page

USE CASE NAME	Admin user Home page
ACTOR	User, Server
DESCRIPTION	Admin user enters the Admin user Home page
Normal Course	<ol style="list-style-type: none">1. User enters the Admin user Home page on the PAS2. System displays the options for the Admin user to upload the Emissions data3. System displays the options for the Admin user to upload the Air Quality Index data
PRECONDITION	User has opened the PAS website
POSTCONDITION	User successfully enters Admin user Home page on the PAS and decides if the user wants to upload Emissions data or Air Quality Index data
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A2. Use case to describe the upload functionality for the Air Quality Index data on the Admin user Home page

USE CASE NAME	Admin user uploads Air Quality Index data
ACTOR	User, Server
DESCRIPTION	Admin user enters the Admin user Home page, selects an Air Quality Index data file and uploads the data in the PAS
Normal Course	<ol style="list-style-type: none"> 1. User enters the Admin user Home page on the PAS 2. System displays the Choose File button for Air Quality Index data 3. User clicks on the Choose File button and selects Air Quality Index data excel or .csv file 4. User clicks on the Upload button
PRECONDITION	User has successfully arrived on the Admin user Home page on the PAS
POSTCONDITION	User successfully uploads Air Quality Index data on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A3. Use case to describe the upload functionality for the Emissions data on the Admin user Home page

USE CASE NAME	Admin user uploads Emissions data
ACTOR	User, Server
DESCRIPTION	Admin user enters the Admin user Home page, selects an Emissions data file and uploads the data in the PAS
Normal Course	<ol style="list-style-type: none"> 1. User enters the Admin user Home page on the PAS 2. System displays the Choose File button for Emissions data 3. User clicks on the Choose File button and selects Emissions data excel or .csv file 4. User clicks on the Upload button
PRECONDITION	User has successfully arrived on the Admin user Home page on the PAS
POSTCONDITION	User successfully uploads Emissions data on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A4. Use case to describe the User Home page

USE CASE NAME	User Home page
ACTOR	User, Server
DESCRIPTION	Regular user enters the User Home page
Normal Course	<ol style="list-style-type: none">1. User enters the User Home page on the PAS2. System displays the year dropdown menu option for the user to select and visualize the Emissions data3. System displays the hyperlink for the user to view Motion Chart for Air Quality Index data4. System displays the hyperlink for the user to view Table for Air Quality Index data5. System displays the hyperlink for the user to view Geo Map for Air Quality Index data
PRECONDITION	User has opened the PAS website
POSTCONDITION	User successfully enters User Home page on the PAS and decides if the user wants to select the year from the year dropdown to visualize Emissions data or clicks on the hyperlinks to view the Motion Chart, Table or Geo Map for Air Quality Index data
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A5. Use case to describe the Visual Representation of Emissions data page

USE CASE NAME	Visual Representation of Emissions data
ACTOR	User, Server
DESCRIPTION	User enters the Visual Representation of Emissions data page, selects a source of pollution from the Emissions Category Tier dropdown menu option to visualize the Emissions data
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User selects a year from the year dropdown menu options and clicks Visualize button 3. User is redirected to the Visual Representation of Emissions data display page 4. User selects a source of pollution from the Emissions Category Tier dropdown menu option to visualize the Emissions data for the selected year
PRECONDITION	User has opened the PAS website
POSTCONDITION	User successfully selects a year from the dropdown menu option on the User home page and visualizes the data on the Visual Representation of Emissions data page for the selected year and the Emissions Category Tier dropdown selection value
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A6. Use case to describe More Visualizations for Emissions data page

USE CASE NAME	More Visualizations for Emissions data
ACTOR	User, Server
DESCRIPTION	User clicks on the hyperlink on the Visual Representation of Emissions data display, arrives at the More Visualizations for Emissions data display and visualizes the data on the basis of state selection from the state dropdown menu and the source of pollution selection from the Category Tier dropdown menu option
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User selects a year from the year dropdown menu options and clicks Visualize button 3. User is redirected to the Visual Representation of Emissions data display page 4. User clicks the hyperlink and is redirected to More Visualizations for Emissions data page 5. User enters the More Visualizations for Emissions data page on the PAS 6. System displays the State dropdown and Category Tier dropdown menu options 7. User selects a State from the State dropdown menu option 8. User selects a source of pollution from the Category Tier dropdown menu option 9. User clicks on Generate Visualization button to visualize the Emissions data
PRECONDITION	User has opened the PAS website and has entered the More Visualizations for Emissions data page upon clicking the hyperlink on the Visual Representation of Emissions data page
POSTCONDITION	User successfully selects the state and the source of pollution from the State and Category Tier dropdown menu options respectively and visualizes the Emissions data on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A7. Use case to describe the Correlation of Emissions data page

USE CASE NAME	Correlation of Emissions data
ACTOR	User, Server
DESCRIPTION	User clicks on the hyperlink on the More Visualizations for Emissions data display, arrives at the Correlation of Emissions data display and visualizes the correlation data on the basis of state 1 and its corresponding year selection and state 2 and its corresponding year selection from the dropdown menu options
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User selects a year from the year dropdown menu options and clicks Visualize button 3. User is redirected to the Visual Representation of Emissions data display page 4. User clicks the hyperlink and is redirected to More Visualizations for Emissions data page 5. User clicks the hyperlink and is redirected to Correlation of Emissions data page 6. User enters the Correlation of Emissions data page on the PAS 7. System displays the State 1 dropdown menu option along with its corresponding year dropdown menu option for the user 8. System displays the State 2 dropdown menu option along with its corresponding year dropdown menu option for the user 9. User selects a state from state 1 dropdown menu option and a year from year dropdown menu option 10. User selects a state from state 2 dropdown menu option and a year from year dropdown menu option 11. User clicks on Calculate Correlation button to visualize the correlation data between two states for same/different year or same state for different years
PRECONDITION	User has opened the PAS website and has entered the Correlation of Emissions data page upon clicking the hyperlink on the More Visualizations for Emissions data page

Table A7. Use case to describe the Correlation of Emissions data page (continued)

POSTCONDITION	User successfully selects the state1 and its corresponding year from the state 1 and year dropdown menu options respectively and state2 and its corresponding year from the state 2 and year dropdown menu options respectively and visualizes the Correlation of Emissions data on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A8. Use case to describe the Motion Chart for Air Quality Index data page

USE CASE NAME	Motion Chart for Air Quality Index data
ACTOR	User, Server
DESCRIPTION	User enters the Motion Chart for Air Quality Index data page upon clicking the hyperlink on the User Home page and visualizes the Air Quality Index data in the form of Motion Chart on the PAS
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User clicks the hyperlink and is redirected to the Motion Chart for Air Quality Index page 3. User enters the Motion Chart for Air Quality Index page on the PAS 4. System displays the Air Quality Index data in the form of Motion Chart
PRECONDITION	User has opened the PAS website and has entered the Motion Chart for Air Quality Index page
POSTCONDITION	User successfully visualizes the Air Quality Index data in the form of Motion Chart on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A9. Use case to describe the Table for Air Quality Index data page

USE CASE NAME	Table for Air Quality Index data
ACTOR	User, Server
DESCRIPTION	User enters the Table for Air Quality Index data page upon clicking the hyperlink on the User Home page and visualizes the Air Quality Index data in the tabular form for a selected year on the PAS
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User clicks the hyperlink and is redirected to the Table for Air Quality Index page 3. User enters the Table for Air Quality Index page on the PAS 4. User selects a year from the year dropdown menu option 5. System displays the Air Quality Index data in the tabular form
PRECONDITION	User has opened the PAS website and has entered the Table for Air Quality Index page
POSTCONDITION	User successfully visualizes the Air Quality Index data in the tabular form for a selected year on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application

Table A10. Use case to describe the Geo Map for Air Quality Index data page

USE CASE NAME	Geo Map for Air Quality Index data
ACTOR	User, Server
DESCRIPTION	User enters the Geo Map for Air Quality Index data page upon clicking the hyperlink on the User Home page and visualizes the Air Quality Index data in the form of Geo Map on the PAS
Normal Course	<ol style="list-style-type: none"> 1. User enters the Home page on the PAS 2. User clicks the hyperlink and is redirected to the Geo Map for Air Quality Index page 3. User enters the Geo Map for Air Quality Index page on the PAS 4. System displays the Air Quality Index data in the form of Geo Map
PRECONDITION	User has opened the PAS website and has entered the Geo Map for Air Quality Index page
POSTCONDITION	User successfully visualizes the Air Quality Index data in the form of Geo Map on the PAS
ASSUMPTIONS	User has entered the PAS website or web-based application