Northeastern University
College of Professional Studies

ALY6010-20698 Module 1 Project

Instructor: Dr. Dee Chiluiza, PhD

Overview and Rationale

This assignment is designed to provide you with hands-on experience in performing descriptive statistical methods on a data set. The data set is provided in an Excel workbook and contains a wide range to data types that you will need to work with.

Assignment Summary

Using the data provided in the attached Excel workbook, apply the methods of graphical and numerical descriptive statistics.

Follow the instructions in the project document to analyze the data presented in the Excel workbook. Then complete a report summarizing the results in your Excel workbook.

For the report, make sure to give a good presentation format to your graphs and tables, don't just copy your raw graphs from excel, make sure your report looks professional.

Important: Explain the benefits of each statistical tool you use, including tables and figures, why you applied those techniques based on the nature of the data, and make substantial observations of the results.

Files

For this project you must submit two files.

- 1) A Power Point file report. Report must be submitted using the link provided.

  Name your file: Week1-Report-LastName-FirstName

- 2) R file. Create a Google folder and share it with your instructor.

  Name your file: Week1-RFile-LastName-FirstName

Part 1: Use the data provided in the Excel file, and perform the following tasks:

Q1. (a) Create a Pareto chart of the top 20 states with the highest population in 2017. **(b**) For these states, which is most populated county in each state? Make a graph or table of your choice to present this data.

Present these two graphs in your report, describe what is the utility of using Pareto charts, and explain your choice of graphs or table for the second question, make meaningful observations of your results.

Q2. Create a frequency and a cumulative frequency histograms for the "International Migration Rate" for **all States**.

In your report, present your graph and describe what was the utility of these graphs in the context of your data.

Q3. Perform numerical descriptive statistics for the "Natural Change" data of all States in the **East coast** and another descriptive for all states in the **West coast**. Include as many parameters as you consider important.

In your report, present a table that allows you to compare the two areas, use the parameter of central tendency and distribution that you consider better describe your data, and explain any difference you observe between the two groups.

Q5. Create a box plot to graph the "International Migration Rate" among **all Counties**. Determine whether there are any outliers, what they are? What is their value? Make observations.

In your report, present your graph, describe the utility box plots and comment on your results.

Q6. Create a scatter plot of the "International Migration Rate" versus "Domestic Migration Rate" of all **states**, and describe whether there exists a correlation or not.

Q7. **Population density** is the relationship of the population per unit area. Using the area of each state in square kilometers, create a graph of your choice that in your opinion, better display this data. Explain your choice of graph and make careful observations of the results. Notice that the area of each state is not given in the excel file, you must find it and provide a reference source.

Grading

Your report values 100 points.

R file values 20 points (all R files will be graded at the end of the course).

Format & Guidelines

A report submitted in Power Point, **no longer than 10 slides**.

Present all relevant graphs showing your work and results, make meaningful observations.

It should include all your findings along with important statistical calculations.

The report should follow the following format:

(i) Introduction slide (Develop ideas and concepts about US immigration and the statistical and graphical tools to be utilized. Be substantial and include references).

(ii) Results slides (Add each task, Q1-Q7, in separate slides. Include all figures and tables with their corresponding titles and descriptive, indicate the main observation and utility of the statistical and graphical tool you used).

(iii) Conclusions slide (Include your conclusions about the observed outcomes, what do you learnt from the statistical tools used and the results? Use references to support your conclusions).

(iv) References slide (Minimum two academic references must be used, proper format will be graded)

Northeastern University
College *of* Professional Studies

Create a good introduction:

Make a list of the topics you must cover in your report (this is a good practice for any report). In this case, for example: (1) US population and migration rates, (2) descriptive statistics, (3) graphical display of data, (4) the purpose of this assignment, (5) what other topics you can think of?

Make sure that all the topics are addressed in your introduction, use academic references when needed. Be brief.

Please remember: your reports are very important, in class or at work. Make them look professional, make them as short as possible but containing all the relevant information. Let me know you have learnt and understood the statistical tools you just used, and for extra points, use deep critical thinking to provide examples of practical applications in real life scenarios.

Northeastern University
College of Professional Studies

Rubric

Category	Above Standard	Meets Standards	Approaching Standards	Below Standards	Not Evident
Excel (or R): Problem Modeling & Set-up	Thoroughly and concisely modeled the problem in Excel (or R) for each method	Accurately modeled the problem in Excel (or R) for each method	Satisfactorily modeled the problem in Excel (or R) for each method.	Partially modeled the problem in Excel (or R) for each method, but there are some gaps in the problem modeling and setup	Did not submit or incompletely modeled the problem in Excel (or R)
Excel (or R): Problem Solution & Accuracy	Thoroughly and efficiently obtained correct and accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Thoroughly obtained accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Satisfactorily obtained correct solutions in Excel (or R) by using the appropriate analytic tools of the software	Partially obtained accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Did not submit or did not obtain accurate solutions in Excel (or R) using the appropriate analytic tools of the software
Word/Report: Problem Description & Introduction	Thoroughly provided a summary of the problem descriptions and introduced the problem using rich and significant ideas	Thoroughly provided a summary of the problem descriptions and problem introduction	Satisfactorily provided a summary of the problem descriptions and problem introduction	Partially provided a summary of the problem descriptions and problem introduction	Did not submit or did not provide a summary of the problem descriptions and problem introduction
Word/Report: Description of Problem Analysis	Thoroughly and accurately described the analytic concepts and theories used in analyzing the problem	Accurately described the analytic concepts and theories used in analyzing the problem	Satisfactorily described the analytic concepts and theories used in analyzing the problem	Partially described the analytic concepts and theories used in analyzing the problem	Did not submit or did not provide a summary of the problem descriptions and problem introduction
Word/Report: Description of Conclusions	Thoroughly described the conclusions and results obtained in the project using a high level of critical thinking and reasoning	Thoroughly described the conclusions and results obtained in the project	Satisfactorily described the conclusions and results obtained in the project	Partially described the conclusions and results obtained in the project	Did not submit or did not describe the conclusions and results obtained in the project
Word/Report: Writing Mechanics, Title Page, & References	Completely free of errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains a minimum of 1000 words	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains a minimum of 1000 words	There are very few errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains a minimum of 1000 words	There are more than five errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are incomplete; or the report contains less than 1000 words	Did not submit; or there are many errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are totally incomplete; or the report contains very few words