Data Set Title
Exploratory Analysis

Garrett Ward, gward2@bellarmine.edu

I.   INTRODUCTION

This data set is about Longley's economic regression data. This data covers multiple variables that contributed to the regression and shows how each played a part in that. I chose this dataset because as a finance and accounting major I am very interested in economics and how certain factors had effects that led to economic regression. Seeing the strong correlation between each factor was very interesting and showed how these led to the economic downturn in Longley.

II.   DATA SET DESCRIPTION

Narrative summary of the data set: e.g. this data set contains 398 samples with 7 columns with various data types. A complete listing is shown in **Table 1**. For data types you want to indicate two things (nominal, ordinal, interval, or ratio) and the Pandas data type. For example, age might be ratio/int32. For missing data, indicate what percentage of data from that column are missing. Ensure you check to for NaN, NA, or any other indicators that actually mean missing data.

Table 1: Data Types and Missing Data

Variable Name	Data Type	Missing Data (%)
GNP.deflator	num	0%
GNP	num	0%
Unemployed	num	0%
Armed.Forces	num	0%
Population	num	0%
Year	int	0%
Employed	num	0%

III.   Data Set Summary Statistics

Narrative introduction to the section.

Table 2: Summary Statistics for XXX (name of dataset)

Variable Name	Count	Mean	Standard Deviation	Min	25th	50th	75th	Max
GNP.deflator	16	101.68	10.79155	83	94.53	100.6	111.25	116.9
GNP	16	384.7	99.39494	234.3	317.9	381.4	454.1	554.9
Unemployed	16	319.3	93.44642	187	234.8	314.4	384.2	480.6
Armed.Forces	16	260.7	69.59196	145.6	229.8	271.8	306.1	359.4
Population	16	117.4	6.956102	107.6	111.8	116.8	122.3	130.1
Year	16	1954	4.760952	1947	1951	1954	1958	1962
Employed	16	65.32	3.511968	60.17	62.71	65.50	68.29	70.55

There should be a table for **EACH** categorical variable.

Table 3: Proportions for XXX (n=yyy)

Category	Frequency	Proportion (%)

After you summarize the categorical variables, generate a correlation matrix for all continuous variables (not categorical – this doesn't make sense)

Table 4: Correlation Table/Tables

	Employed	Year	Population	Armed.Forces	Unemployed	GNP	GNP.deflator	
Employed	1	0.97	0.96	0.46	0.5	0.98	0.97	
Year	0.97	1	0.99	0.42	0.67	1	0.99	
Population	0.96	0.99	1	0.36	0.69	0.99	0.98	
Armed.Forces	0.46	0.42	0.36	1	-0.18	0.45	0.46	
Unemployed	0.5	0.67	0.69	-0.18	1	0.6	0.62	
GNP	0.98	1	0.99	0.45	0.6	1	0.99	
GNP.deflator	0.97	0.99	0.98	0.46	0.62	0.99	1	

After the table with the raw data, include a heatmap of the correlation matrix as a figure.

IV.   DATA SET GRAPHICAL EXPLORATION

Narrative introduction to the section. In each section below, indicate any interesting distributions, anomalies, imbalance, etc. that you notice.
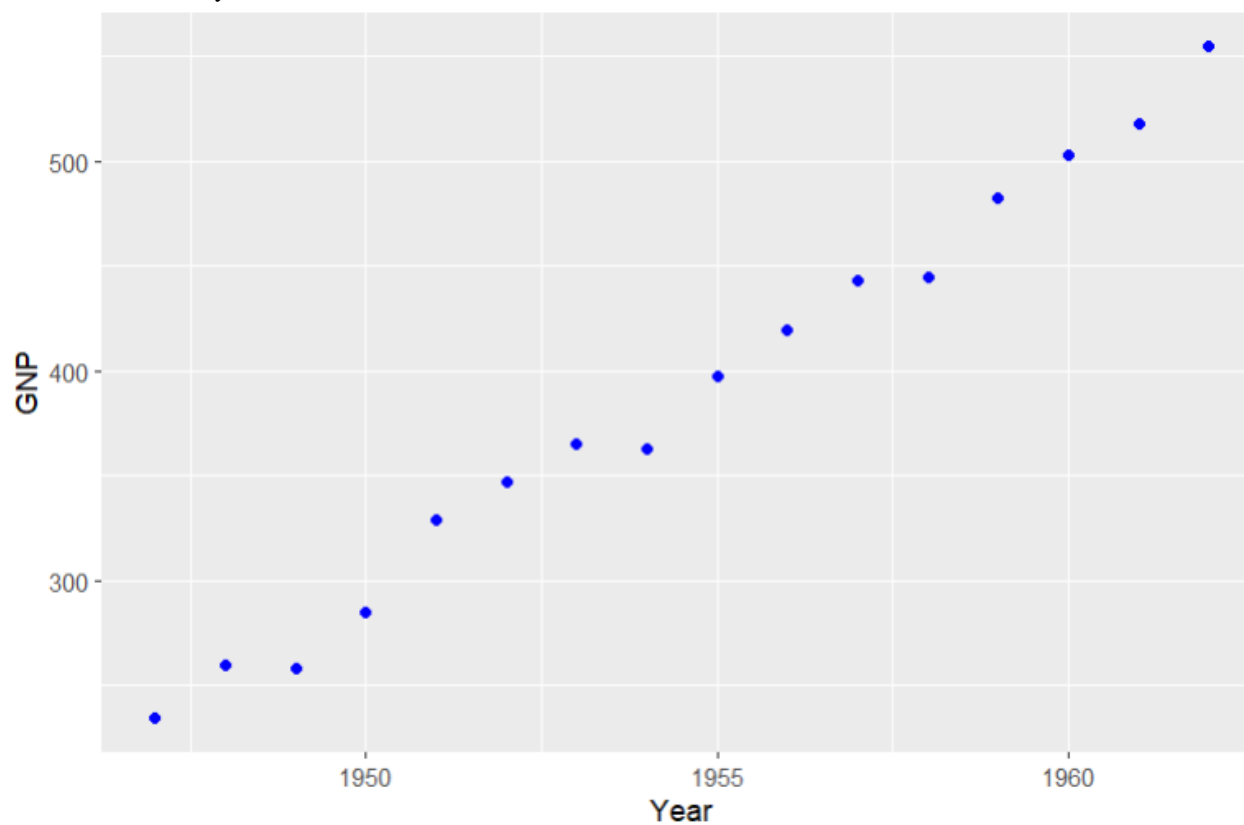

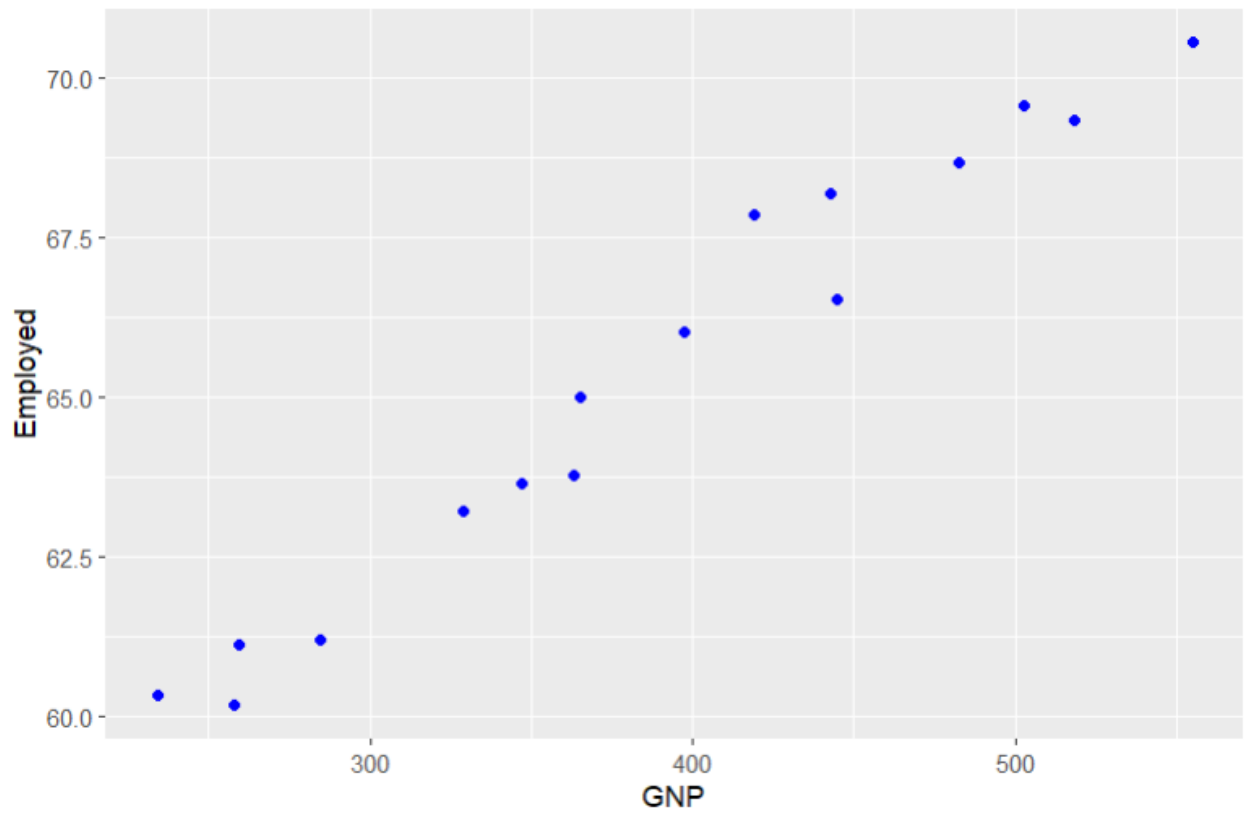*Figure 1 - Scatter plot of national GNP by year*

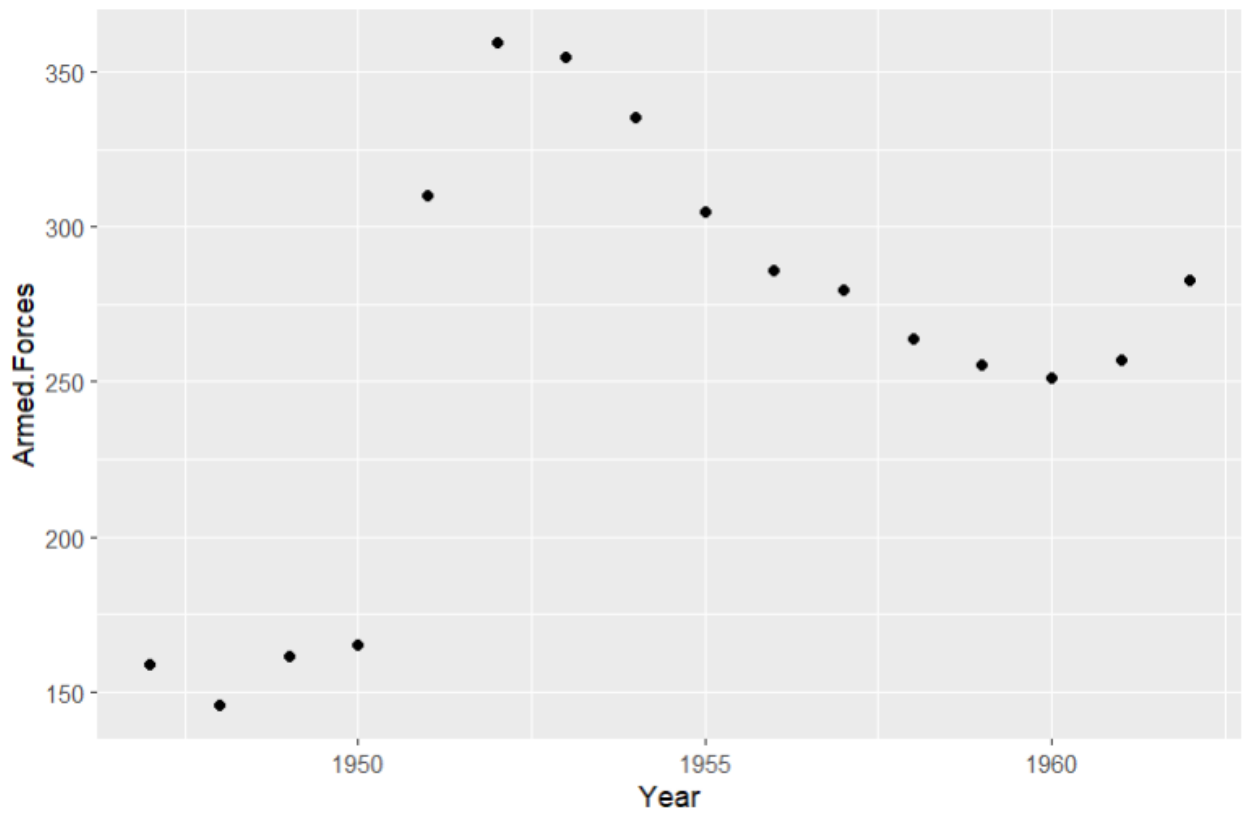*Figure 2 - Scatter plot comparing the Employment level and national GNP*

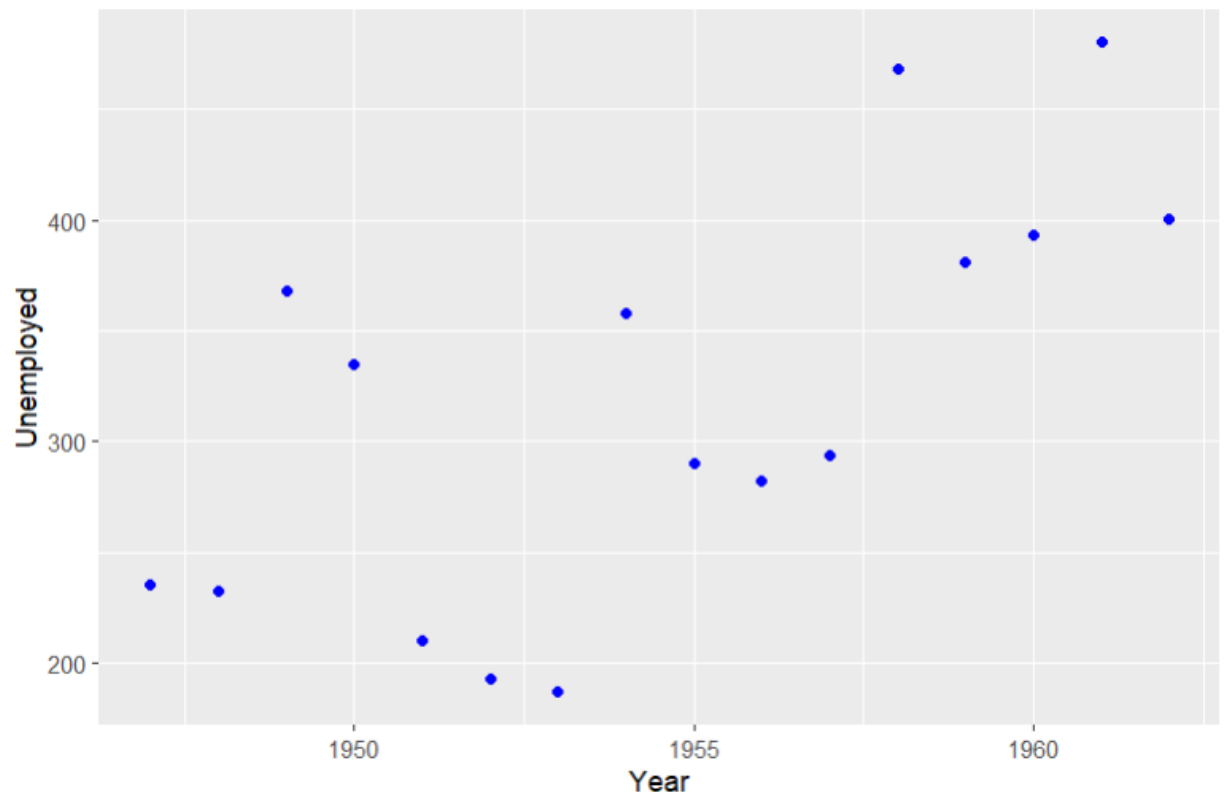*Figure 3 - Scatter plot of amount of Armed Forces each year*
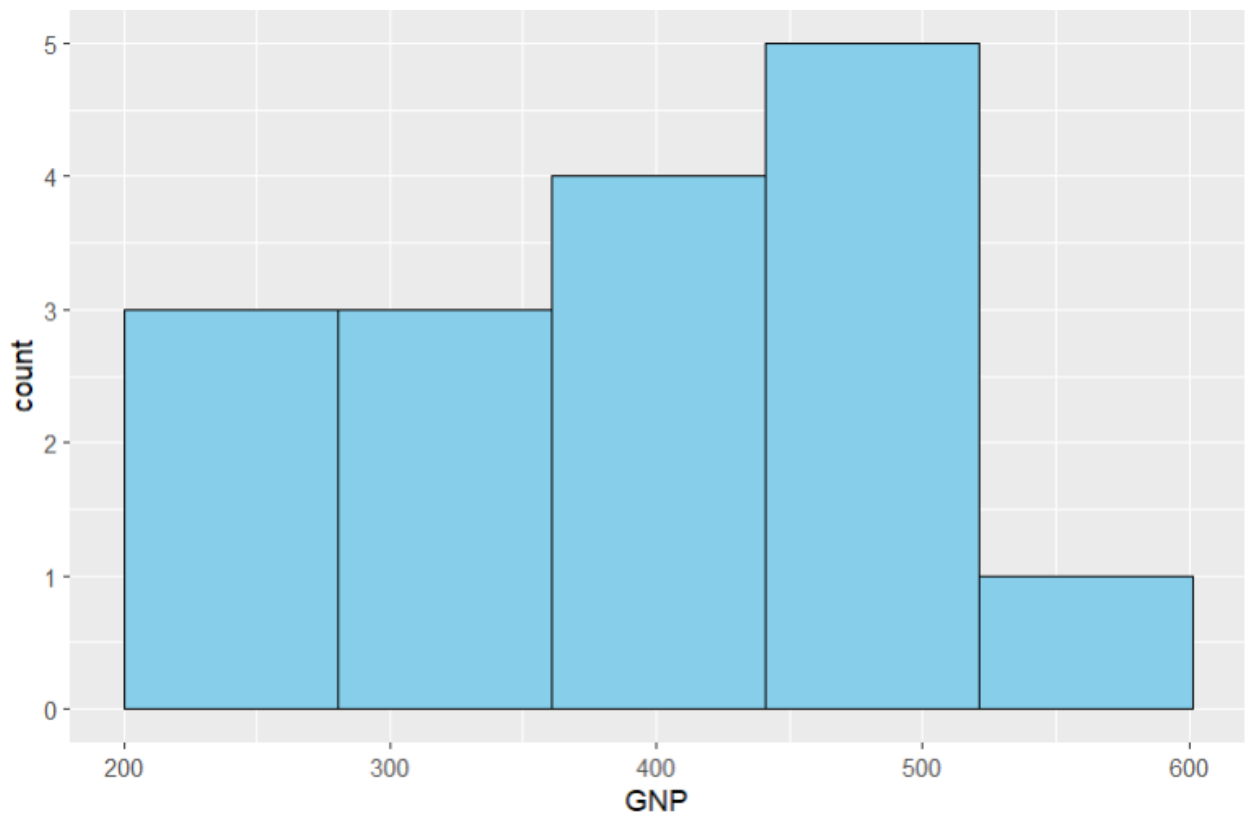
Figure 4 - Scatter plot of the unemployed total by year

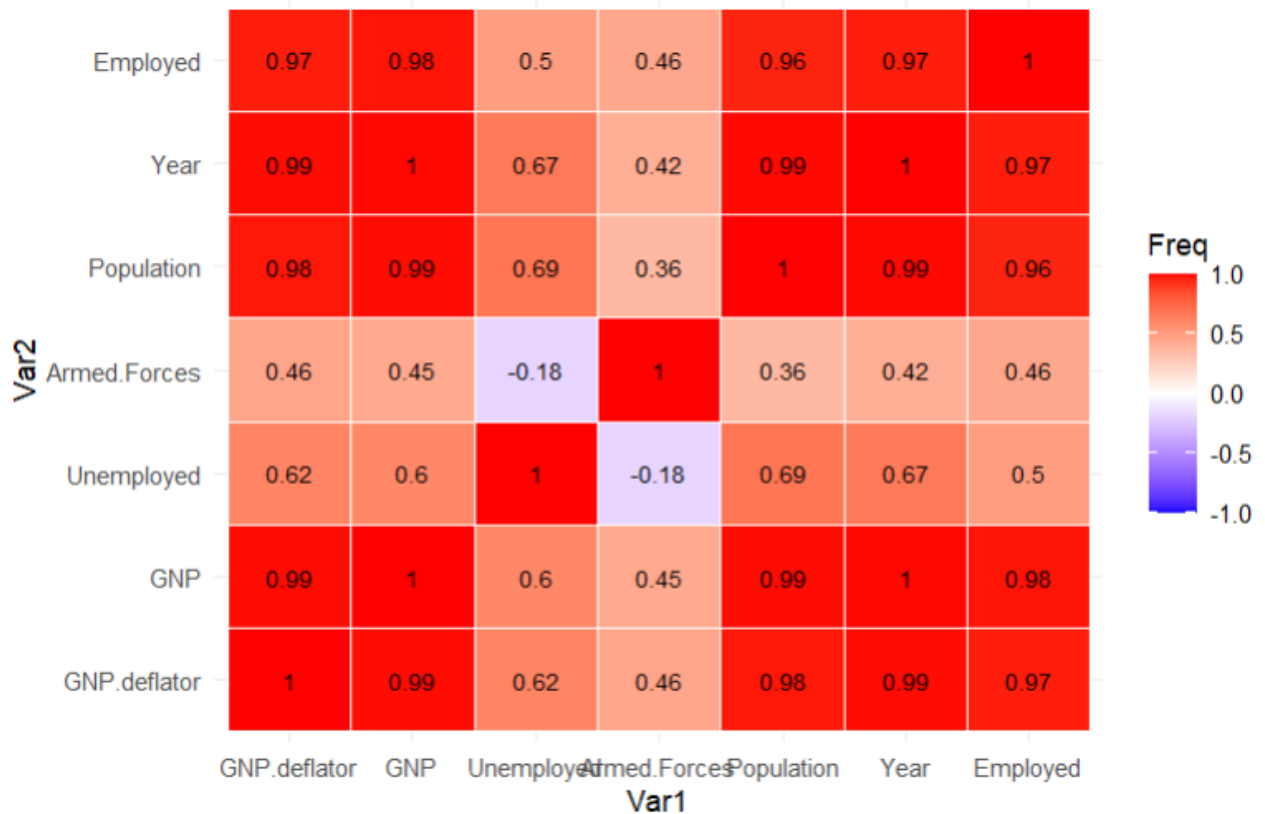*Figure 5 - Histogram to show the count of GNP levels*

Figure 6 - Correlation matrix of the variables

V. SUMMARY OF FINDINGS

After conducting the research, I found that there was very similar growth year over year from GNP and Employment levels each year. There was also a steady increase in the unemployment levels which contributed to the regression of the economy. The correlation map showed that there were very strong, positive relationships between all the variables in the dataset except for the correlation of Armed Forces and unemployment. Finally, the most common GNP levels were between 450 and 525. However, the lower GNP levels made up for a majority of the data than the larger GNP levels which can be attributed to the eventual economic regression. Overall, this dataset was very interesting and showed how certain factors that can be present today have affected previous economy's and what they led to.