POLS/COMS228 – Data Visualisation

HW 3

You will submit your homework as 2 files: 1) a R Markdown (.Rmd) file and 3) an HTML file knitted from your R Markdown file. You must upload both files to AKO|LEARN using Xxx.Rmd and Xxx.html as file names where Xxx is your last name.

I will review both your files. To receive full credit:

- You must submit your files on time. They must be named exactly as specified, and your .Rmd file must knit without errors to produce a .html file. Double check your file for errors before submitting.
- The .html file should read as a well written report with plots and your comments and answers included in the file. Your report should not contain any extra or foreign material, such as leftovers from code or notes.
- The code in your .Rmd files must be clear, readable, and well organized.

1. Plot GDP by industry data

Alice is preparing a report about regional gross domestic products (GDP), in other words, how much each industry in New Zealand contributes to the country's GDP. She is particularly interested in comparing contributions by industries (in NZD) between the South and North Islands in the year 2000. Alice finds that Manufacturing contributes the most to the GDP in both the South and North Islands. She decides to structure the report around the importance of manufacturing in New Zealand's GDP. The report will be presented to a general audience. The source of the data is Stats NZ.

You are tasked with producing a data visual to support the data storytelling of the report. Produce one data visualization that plots how much (in dollars) each industry contributes to the GDP. Differentiate between the North and the South Islands. Emphasize the values for manufacturing industry.

Refine your graph, modifying elements such as the title, subtitle, labels, caption, colours, axes. Make the graph visually appealing. Follow guidelines for better data visualisations from your reading and lectures.

Load "gdp_by_industry.csv" file as gdp. The dataset provides you GDP by industry and year for New Zealand regions, islands, and total for the country in NZD. You will work with data for year 2000.

After loading the file run the following code to clean the dataset and keep the GDP by industry data for North and South islands for 2000.

```
gdp <- read_csv("gdp_by_industry.csv")
```

```
gdp_2020 <- subset(gdp, Period == 2000.03 & Group == "Gross domestic
product, by region and industry" & (Series_title_2 == "Total North Island"
| Series_title_2 == "Total South Island"))
```

```
gdp_plot <- select(gdp_2020, c("Data_value", "UNITS", "Series_title_3",
"Series_title_2"))
```

```
You will notice that the cleaned data includes some redundant categories
and well as total categories. We should remove them. Run the following
code:
```

```
gdp_plot <- gdp_plot %>%

      filter(!row_number() %in% c(3, 4, 5, 23, 24, 25, 28, 29, 30, 48, 49,
50))
```

You should now have a dataframe named gdp_plot in your environment. By using gdp_plot, plot the data.

Codebook:

Data_value – GDP in dollars

UNITS – unit, dollars

Series_title_3 – Industry

Series_title_2 – denotes North or South Island

You will be graded based on the choice of the graph (appropriateness for the data), accuracy of data representation (data plotted accurately), and visual appeal of the graph (refined graph).

Report all your work.