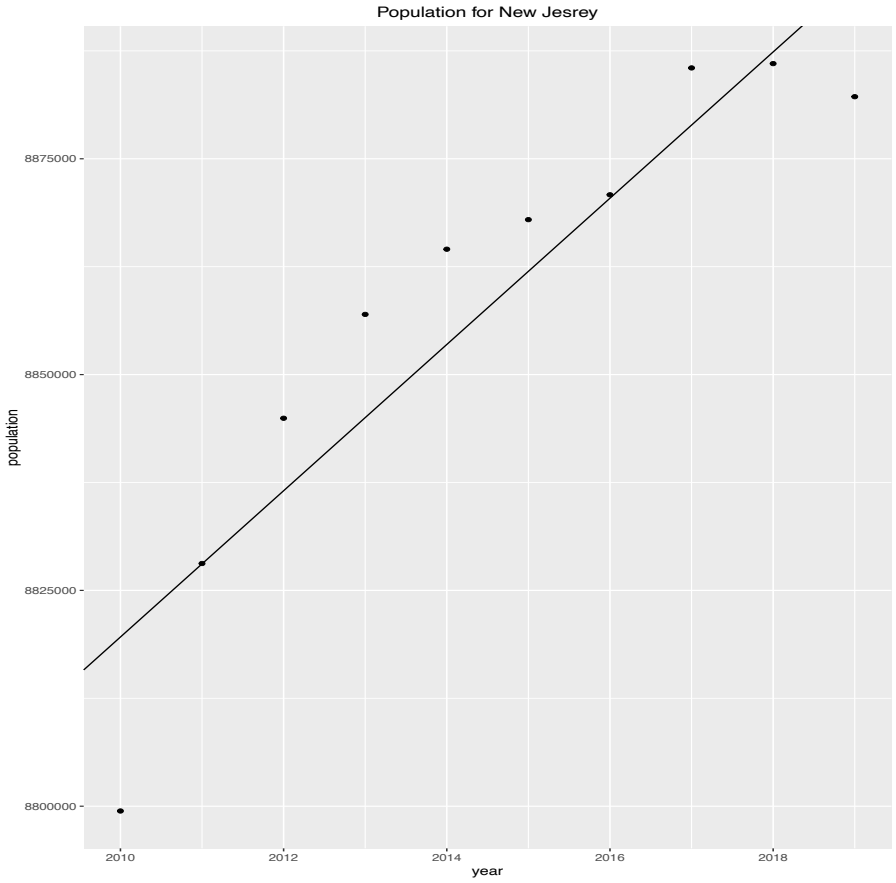
C997 Task Prompt Responses

Elizabeth Sweet

Western Governors University

C997 Task Prompt Responses

A. Create a linear regression analysis with R to predict the size of the population for the state you live in based on the "Current Estimated Data." Provide a screenshot of your results.



I used stock overflow to learn how to center the title of the graph (Center Plot title in ggplot2, 2017). I also used help with making the tick marks integers (How to display only integer values on an axis using ggplot2, 2013).

B. Explain how you prepared the data from part A and how the dataset was imported into R, including a screenshot of your results.

I downloaded the required file from the Census Bureau website (Population, Population Change, and Estimated Components of Population Change: April 1, 2020 to July 1, 2019, 2020). I imported it into R using read.csv. Next, I converted it to a data table. I isolated the row for New Jersey and removed the columns that were not needed. I then changed the column named from POPESTIMATE20** to the year. Using gather, I changed the table from wide to long form. Finally, I changed the year variable to integers.

```
data <- read.csv('~/Desktop/nst-est2019-alldata.csv')
dt <- data.table(data)
njpop <-  dt[36,]
njpop2 <- njpop[,1:17]
njpop3 <-  njpop2[,-(1:7)]
njpop3 <- rename(njpop3, c('2010' = 'POPESTIMATE2010', '2011' = 'POPESTIMATE2011',
'2012' = 'POPESTIMATE2012', '2013' = 'POPESTIMATE2013', '2014' = 'POPESTIMATE2014',
'2015' = 'POPESTIMATE2015', '2016' = 'POPESTIMATE2016', '2017' = 'POPESTIMATE2017',
'2018' = 'POPESTIMATE2018', '2019' = 'POPESTIMATE2019'))
njpop4 <- gather(njpop3, 'year', 'population')
njpop4$year <- as.integer(njpop4$year)
```

Code to prepare data

	year	population
1	2010	8799446
2	2011	8828117
3	2012	8844942
4	2013	8856972
5	2014	8864525
6	2015	8867949
7	2016	8870827
8	2017	8885525
9	2018	8886025
10	2019	8882190

Resulting table

C. Create an R script that will tabulate a statistical description of the model using R's summary() function and provide a screenshot of your results.

```
Call:
lm(formula = population ~ year, data = njpop4)

Residuals:
   Min     1Q Median     3Q    Max
-21080  -1931   2264   7026  11029

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -8209120    2481036  -3.309 0.010718
year            8472       1232   6.879 0.000127

(Intercept) *
year        ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11190 on 8 degrees of freedom
Multiple R-squared:  0.8554,    Adjusted R-squared:  0.8373
F-statistic: 47.32 on 1 and 8 DF,  p-value: 0.0001272
```

Screenshot of statistical summary

D. **Predict the population size of your state in five years using a linear regression form part A and provide a screenshot of your results.**

Using information found on Stack Overflow, I predicted that the population of New Jersey would be 8,947,613 (Predict() - Maybe I'm not understanding it, 2012).

```
> future <- data.frame(year = c(2025))
> View(future)
> predict(regression, future)
      1
8947613
>
```

References

Center Plot title in ggplot2. (2017). Retrieved from Stock Overflow: https://stackoverflow.com/questions/40675778/center-plot-title-in-ggplot2

How to display only integer values on an axis using ggplot2. (2013). Retrieved from Stack Overflow: https://stackoverflow.com/questions/15622001/how-to-display-only-integer-values-on-an-axis-using-ggplot2

Population, Population Change, and Estimated Components of Population Change: April 1, 2020 to July 1, 2019. (2020). Retrieved from United States Census Bureau: https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-cities-and-towns.html

Predict() - Maybe I'm not understanding it. (2012). Retrieved from Stack Overflow: https://stackoverflow.com/questions/9028662/predict-maybe-im-not-understanding-it