Part II: R

I. Create a linear regression analysis with R to predict the size of the population for the state you live in for 2020 based on the "Current Estimates Data" dataset (see weblink below). Provide a screenshot of your results [2].

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
#Confidence Interval for the Utah population estimate for the year 2020
y_pred <- lm(regressor, data = test_set)
new_data <-data.frame(Year=2020)
utah_predict<- predict(y_pred,new_data,interval="predict",level=.95)
utah_predict
```
```

```
       fit       lwr       upr
 1 3254141 3002358 3505925
```

```
Using a 95% confidence interval, the results of the linear regression model
indicate the estimated population for the state of Utah will be between
3,002,358 and 3,505,925 people.
```

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
#Utah state population estimate for the year 2020
utah_predict<- predict(y_pred,new_data)
utah_predict
```
```

```
       1
 3254141
```

```
The Linear Regression model result for the 2020 Utah state population is
3,254,141 people.
```

J. Explain how you prepared the data from part I and how the dataset was imported into R, including screenshots of your results.

Step 1: Downloaded the "Current Estimates" dataset as a CSV file. Using MS Excel, I deleted all unnecessary columns, keeping only the Utah census data from 2010 – 2017. I then saved the cleaned dataset as a new CSV file named ut_census.

Step 2: Using the Readr library, I imported the cleaned Utah Census Bureau data into R.

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
# Load the Data
UT_data<- read_csv("~/Desktop/MSDA Portfolio/2. DS Tools &
Techniques/P2. R/ut_census.csv")

View(UT_data)
```

	Year	Pop_Est
1	2010	2775260
2	2011	2815430
3	2012	2854222
4	2013	2899961
5	2014	2938671
6	2015	2984917
7	2016	3044321
8	2017	3101833

K. Create an R script that will tabulate a statistical description of the estimated 2020 data using R's summary() function and provide a screenshot of your results. [2]

```r
```{r message=FALSE, warning=FALSE}
# Split the dataset into the Training set and Test set
set.seed(123)
split = sample.split(UT_data$Pop_Est, SplitRatio = 2/3)
training_set = subset(UT_data, split == TRUE)
test_set = subset(UT_data, split == FALSE)

# Fit the Simple Linear Regression to the Training set
regressor = lm(formula = Pop_Est ~ Year,
               data = training_set)
```
```

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
summary(regressor)
```
```

```
Call:
lm(formula = Pop_Est ~ Year, data = training_set)

Residuals:
      1        2        3        4        5
 4420.4    308.1  -5182.2  -7334.0   7787.7

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -86236535    2847644  -30.28 7.91e-05 ***
Year            44282       1415   31.30 7.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7324 on 3 degrees of freedom
Multiple R-squared:  0.9969,    Adjusted R-squared:  0.9959
F-statistic: 979.7 on 1 and 3 DF,  p-value: 7.165e-05
```

L. Predict the population size of your state in five years using a linear regression from part I and provide a screenshot of your results [2].

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
#Confidence Interval for the Utah population estimate for the year 2023
y_pred <- lm(regressor, data = test_set)
new_data <-data.frame(Year=2023)
utah_predict<- predict(y_pred,new_data,interval="predict",level=.95)
utah_predict
```
```

```
      fit      lwr      upr
1 3408259 3049965 3766552
```

Using a 95% confidence interval, the results of the linear regression model
indicate the estimated population for the state of Utah will be between
3,049,965 and 3,766,552 people.

```r
```{r echo=FALSE, message=FALSE, warning=FALSE}
#Utah state population estimate for the year 2023
utah_predict<- predict(y_pred,new_data)
utah_predict
```
```

```
      1
3408259
```

Predicted Utah state population for the year 2023 is 3,408,259 people.

Resources

1. Dataset - Retrieved August 30, 2018, from https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/state/asrh/

2. Eremenko, K., & De Ponteves, H. (n.d.). Machine Learning A-Z™: Hands-On Python & R In Data Science. Retrieved September 1, 2018, from https://www.udemy.com/