

Stat 519 Project#1 Preliminary Report due 1/28

Monday in class

*Your full Name here + your Github address (e.g. Stephen Lee +
<https://github.com/sauchilee/Stat519>)*

January 18, 2019

Describe your observations (units) and variables of your Data. Input your data set (in .csv format) here. For example:

```
your_data=read.csv(file="https://raw.githubusercontent.com/sauchilee/Stat519/master/Data/iris.csv",  
                    header=T)  
dim(your_data)
```

```
## [1] 150  6
```

```
tail(your_data)
```

```
##      X Sepal.Length Sepal.Width Petal.Length Petal.Width  Species  
## 145 145          6.7          3.3          5.7          2.5 virginica  
## 146 146          6.7          3.0          5.2          2.3 virginica  
## 147 147          6.3          2.5          5.0          1.9 virginica  
## 148 148          6.5          3.0          5.2          2.0 virginica  
## 149 149          6.2          3.4          5.4          2.3 virginica  
## 150 150          5.9          3.0          5.1          1.8 virginica
```

The data set of your project should have well defined experimental units (observations) with at least 10 quantitative (i.e., numerical) variables, plus at least 4 categorical variables. One of the categorical variable should have at least 3 levels/groups with at least 30 observations for each level/group.

State clearly your research objectives and questions clearly.

Perform EDA (Exploratory Data Analysis) by computing summary statistics and presenting visualizations in:

- Enhanced scatterplots, Convex hull, Chi-plot, Bivariate boxplot, Bivariate density estimator, Bubble plot, Scatterplot matrix, 3-D scatterplot, Star plot, Chernoff faces, and
- R Graphics