Exercises week 4

Antoine Vernet

Exercise 1

The States dataset has some information about US states. You can find this dataset in the `car` package.

```
library(car)
st <- States
# This is for convenience and transforms the rownames into an additional column
st[, "States"] <- rownames(st)
# You can access the description of the dataset with the following:
# ?States
```

1. Examine the correlations between the SATV and SATM scores. Can you explain your finding in one or two sentences?

```
# Your code here
```

2. Examine the correlations between the state spending on public education (dollars) and the 2 different SAT scores. Can you explain what you find in one or two sentences?

```
# Your code here
```

3. Create a plot showing how the SATV and SATM scores covary with the average teacher's salary (pay).

```
# Your code here
```

4. Can you build a simple linear regression model trying to explain the SATM score (but not including the SATV score as a variable)? Explain why you include each of the variables in your model: what leads you to think this variable will have an effect on the score (Write no more than 3-5 sentences)?

```
# Your code here
```

Exercise 2

The nuclear dataset in the `boot` package contains information about nuclear power station construction data. This dataset has 32 rows and 11 columns.

```
library(boot)
ncl <- nuclear
# You can access a description of the dataset using ?nuclear
```

Can you build a simple model of the determinants of cost for this dataset? In this case, your goal is to achieve the largest possible value for adjusted R-square.

```
# Your code here
```

Exercise 3

The car package contains a dataset called `Salaries` which gives some information about the faculty in an american college in 2008-2009, including salaries.

```
library(car)
slr <- Salaries
# You can access a description of the dataset using ?Salaries
```

1. Plotting the relationship between gender (the `sex` variable) and wage (the `salary` variable), does it looks likely that there is gender discrimination (1-2 sentences)?

```
# Try using ggplot2 with the geom_boxplot type.
```

2. Redo the previous plot, but this time separating the plot by rank. Does discrimination looks likely (1-2 sentences)?

```
# Try using ggplot2 with the geom_boxplot type.
```

3. Can you build a regression model that examine the effect of gender on salary? When you add additional variables, does the relationship between salary and gender changes? What else influences salary? (Write no more than 3-5 sentences)

```
# Your code here
```

4. Would you recommend that this college take action about its salary structure? Would you take any other actions? (3-5 sentences)