Week3DA

Jonas.

1 Introduction

1.1 Subsection

1.2 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
evals.scores <- evals %>%
  select(score, bty_avg)
head(evals.scores)
```

```
## # A tibble: 6 x 2
##   score bty_avg
##   <dbl>   <dbl>
## 1   4.7       5
## 2   4.1       5
## 3   3.9       5
## 4   4.8       5
## 5   4.6       3
## 6   4.3       3
```

The number of observations in evals is 463‘

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

```
kable(head(iris, n = 5), caption = '\\label{tab:iris} The first 5 rows of the iris data.') %>%
  kable_styling(font_size = 10, latex_options = 'HOLD_position')
```

Table 1: The first 5 rows of the iris data.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

```
iris %>%
  group_by(Species) %>%
  summarise(n=n(), Mean = round(mean(Sepal.Length), digits = 1),
            St.Dev=round(sd(Sepal.Length),digits = 1),
            Min = min(Sepal.Length),
            Q1 = quantile(Sepal.Length, 0.25),
            Median = median(Sepal.Length),
            Q3 = quantile(Sepal.Length,0.75),
            Max = max(Sepal.Length)) %>%
  kable(caption = '\\label{tab:summaries} Summary statistics on the sepal length by species of irises.'
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 2: Summary statistics on the sepal length by species of irises.

Species	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
setosa	50	5.0	0.4	4.3	4.800	5.0	5.2	5.8
versicolor	50	5.9	0.5	4.9	5.600	5.9	6.3	7.0
virginica	50	6.6	0.6	4.9	6.225	6.5	6.9	7.9

```
my_skim <- skim_with(base = sfl(n = length))
iris %>%
  group_by(Species) %>%
  select(Sepal.Length,Species) %>%
  my_skim() %>%
  transmute(Variable=skim_variable, Species=Species, n=n, Mean=numeric.mean, SD=numeric.sd,
            Min=numeric.p0, Median=numeric.p50,  Max=numeric.p100,
            IQR = numeric.p75-numeric.p50) %>%
   kable(caption = '\\label{tab:summaryskim} Summary statistics of the sepal length by species of irises
         booktabs = TRUE, linesep = "", digits = 2) %>%
   kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 3: Summary statistics of the sepal length by species of irises (produced using skim() function).

Variable	Species	n	Mean	SD	Min	Median	Max	IQR
Sepal.Length	setosa	50	5.01	0.35	4.3	5.0	5.8	0.2
Sepal.Length	versicolor	50	5.94	0.52	4.9	5.9	7.0	0.4
Sepal.Length	virginica	50	6.59	0.64	4.9	6.5	7.9	0.4

```
iris %>%
  group_by(Species) %>%
  summarise(n = n(), mean = round(mean(Petal.Width),digits = 1),
            sd = round(sd(Petal.Width),digits = 1),
            min = min(Petal.Width),
            q1 = quantile(Petal.Width,0.25),
            median = median(Petal.Width),
            q3 = quantile(Petal.Width,0.75),
            max = max(Petal.Width)) %>%
  kable(caption = '\\label{tab:summaries2} Summary statistics on the sepal length by species of irises'
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Table 4: Summary statistics on the sepal length by species of irises

Species	n	mean	sd	min	q1	median	q3	max
setosa	50	0.2	0.1	0.1	0.2	0.2	0.3	0.6
versicolor	50	1.3	0.2	1.0	1.2	1.3	1.5	1.8
virginica	50	2.0	0.3	1.4	1.8	2.0	2.3	2.5

```
iris %>%
  mutate(sepal.length.class = if_else(Sepal.Length<5.5,'small','large')) %>%
  group_by(Species, sepal.length.class) %>%
  summarise(n = n()) %>%
  spread(sepal.length.class, n) %>%
  kable() %>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Species	large	small
setosa	5	45
versicolor	44	6
virginica	49	1

```
iris %>%
  mutate(sepal.length.class = if_else(Sepal.Length<5.5,'small','large')) %>%
  group_by(Species, sepal.length.class) %>%
  summarise(n=n()) %>%
  mutate(prop=n/sum(n)) %>%
  kable(digits = 2) %>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Species	sepal.length.class	n	prop
setosa	large	5	0.10
setosa	small	45	0.90
versicolor	large	44	0.88
versicolor	small	6	0.12
virginica	large	49	0.98
virginica	small	1	0.02

```r
iris %>%
  mutate(sepal.length.class = if_else(Sepal.Length<5.5,'small','large')) %>%
  group_by(Species, sepal.length.class) %>%
  summarise(n=n()) %>%
  mutate(prop=n/sum(n)) %>%
  subset(select=c("Species","sepal.length.class","prop")) %>%
  spread(sepal.length.class,prop) %>%
  kable(digits = 2) %>%
  kable_styling(font_size = 10, latex_options = "hold_position")
```

Species	large	small
setosa	0.10	0.90
versicolor	0.88	0.12
virginica	0.98	0.02

```r
model <- lm(Sepal.Length~Species, data = iris)
get_regression_table(model) %>%
  dplyr::select(term,estimate) %>%
  #Note that it seems necessary to include "dplyr::" here!!
  kable(caption = '\\label{tag:reg} Estimate of the parameters from the fitted linear regression model.
  kable_styling(latex_options = "hold_position")
```

Table 5: Estimate of the parameters from the fitted linear regression model.

term	estimate
intercept	5.006
Speciesversicolor	0.930
Speciesvirginica	1.582

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Table 6: The fist 5 rows of the iris data.

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

Table 1 displays the first 5 rows of the iris data

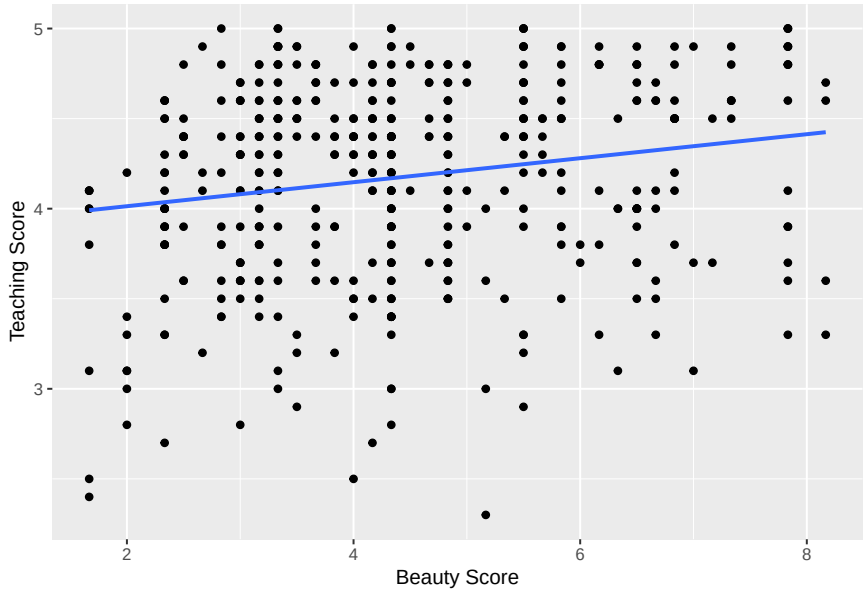
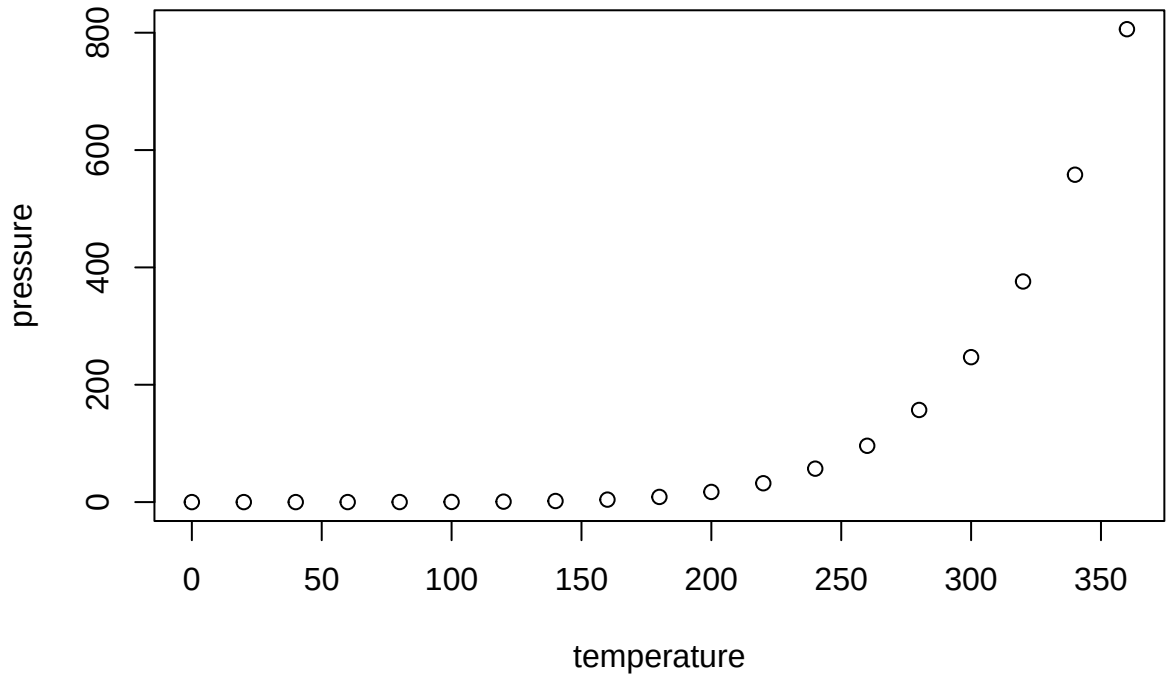


Figure 1: Relationship between teaching and beauty scores. The best-fitting line has been superimposed.

1.3 Including Plots

You can also embed plots, for example:



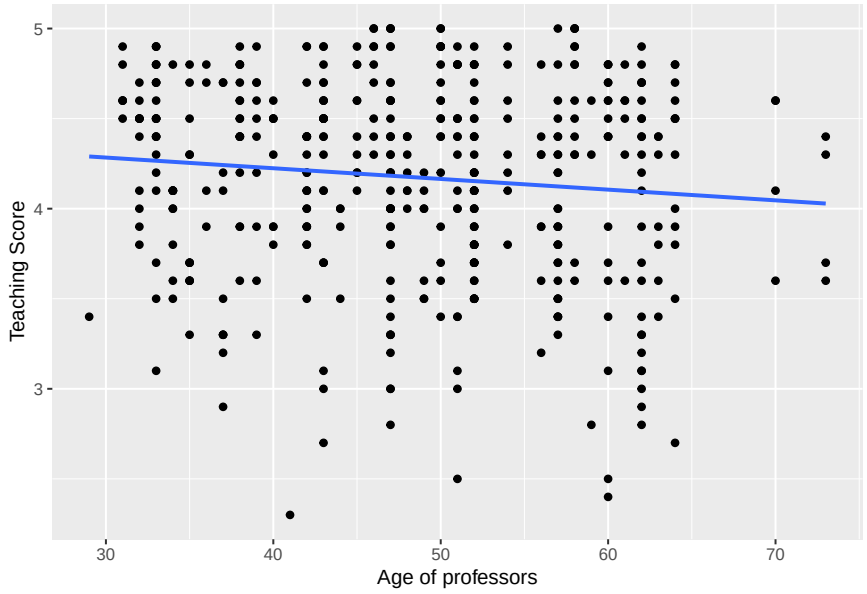


Figure 2: Relationship between the teaching score and the age of the professors

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.