

# SWP1 - Group 6

*Thomas Siskos, Julian Cantor, Oliver Brose, Konstantin Haesler*

*May 3, 2018*

## Questions 1, 2, 3 & 4

There are 397 observation and 6 variables in the dataset, 21 professors with more than 40 yrs of experience and 54 professors with a salary higher than 150,000. On the other hand, the mean salary of the professors with more than 20 yrs of experience is \$122,104.

## Question 5:

To find out more of the dataset we computed some descriptive statistics.

```
summary(Salaries)
```

```
##           rank      discipline yrs.since.phd   yrs.service      sex
## AsstProf : 67   A:181         Min.    : 1.00   Min.    : 0.00   Female: 39
## AssocProf: 64   B:216         1st Qu.:12.00  1st Qu.: 7.00   Male  :358
## Prof      :266                Median :21.00  Median :16.00
##                Mean    :22.31   Mean    :17.61
##                3rd Qu.:32.00  3rd Qu.:27.00
##                Max.    :56.00   Max.    :60.00
##      salary
## Min.    : 57800
## 1st Qu.: 91000
## Median :107300
## Mean    :113706
## 3rd Qu.:134185
## Max.    :231545
```

## Question 6:

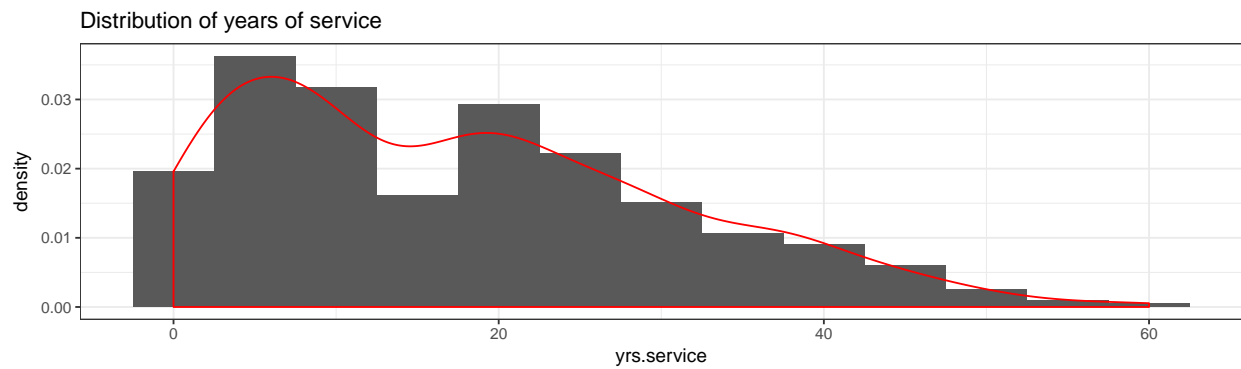
The following chart shows the counts and proportions by rank and sex.

```
counts = Salaries %>%
  count(sex, rank) %>%
  mutate(prop = n / sum(n))
print(counts)
```

```
## # A tibble: 6 x 4
##   sex    rank      n  prop
##   <fct> <fct>   <int> <dbl>
## 1 Female AsstProf    11 0.0277
## 2 Female AssocProf   10 0.0252
## 3 Female Prof       18 0.0453
## 4 Male   AsstProf    56 0.141
## 5 Male   AssocProf   54 0.136
## 6 Male   Prof      248 0.625
```

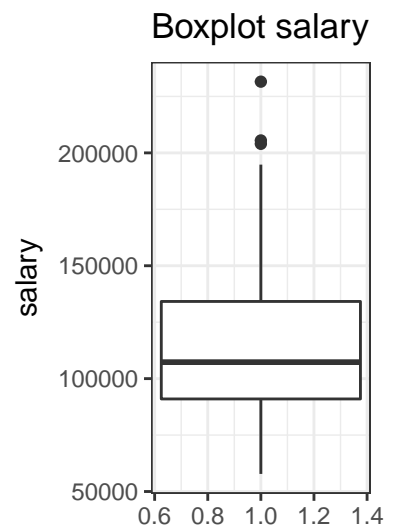
### Question 7:

The histogram below shows the desired output.



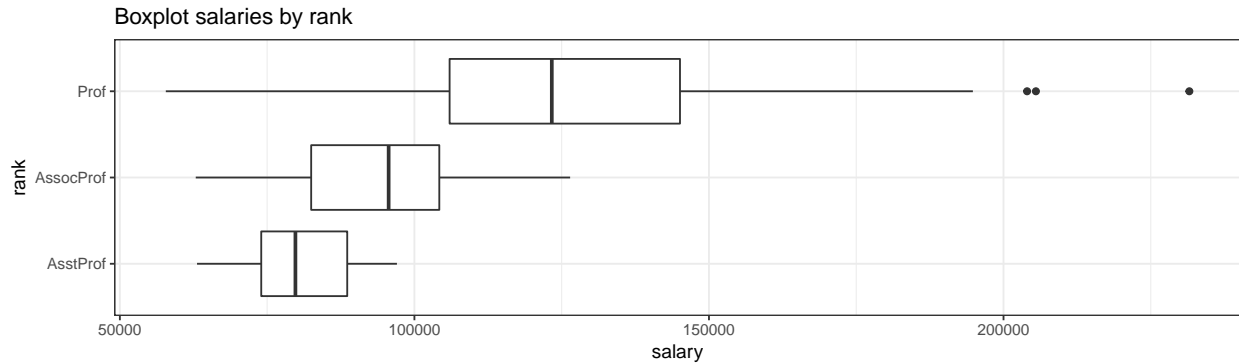
### Question 8:

The bloxplot below shows the desired output.



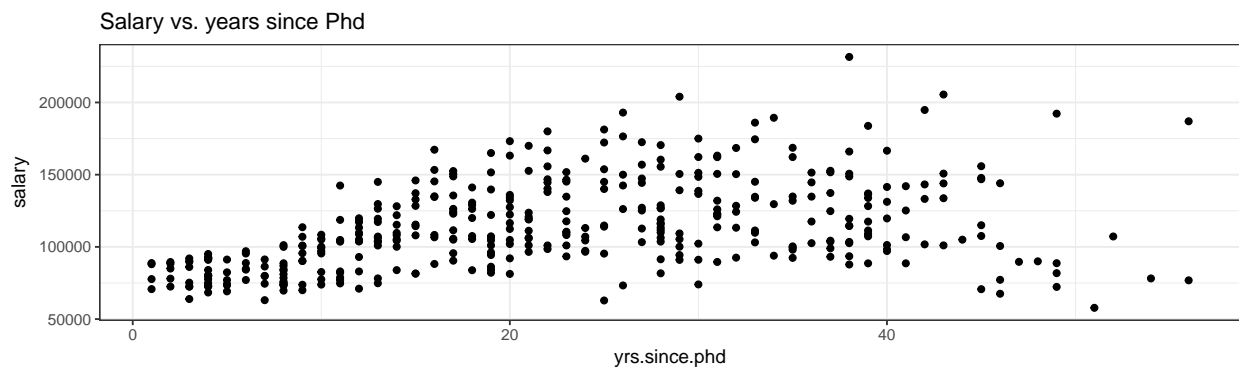
### Question 9:

The bloxplot below shows the desired output.



### Question 10:

The scatterplot shows the desired result.



### Question 11:

The correlation between salary and years since phd is 0.41: With the model output shown below one can see that the p-value is less than 0.05 and thus the correlation has statistical significance.

```
cor(Salaries$salary, Salaries$yrs.since.phd)
```

```
[1] 0.4192311
```

```
cor(Salaries$salary, Salaries$yrs.service)
```

```
[1] 0.3347447
```

```
model = lm("salary ~ yrs.since.phd", data = Salaries)
print(summary(model))
```

Call: `lm(formula = "salary ~ yrs.since.phd", data = Salaries)`

Residuals: Min 1Q Median 3Q Max -84171 -19432 -2858 16086 102383

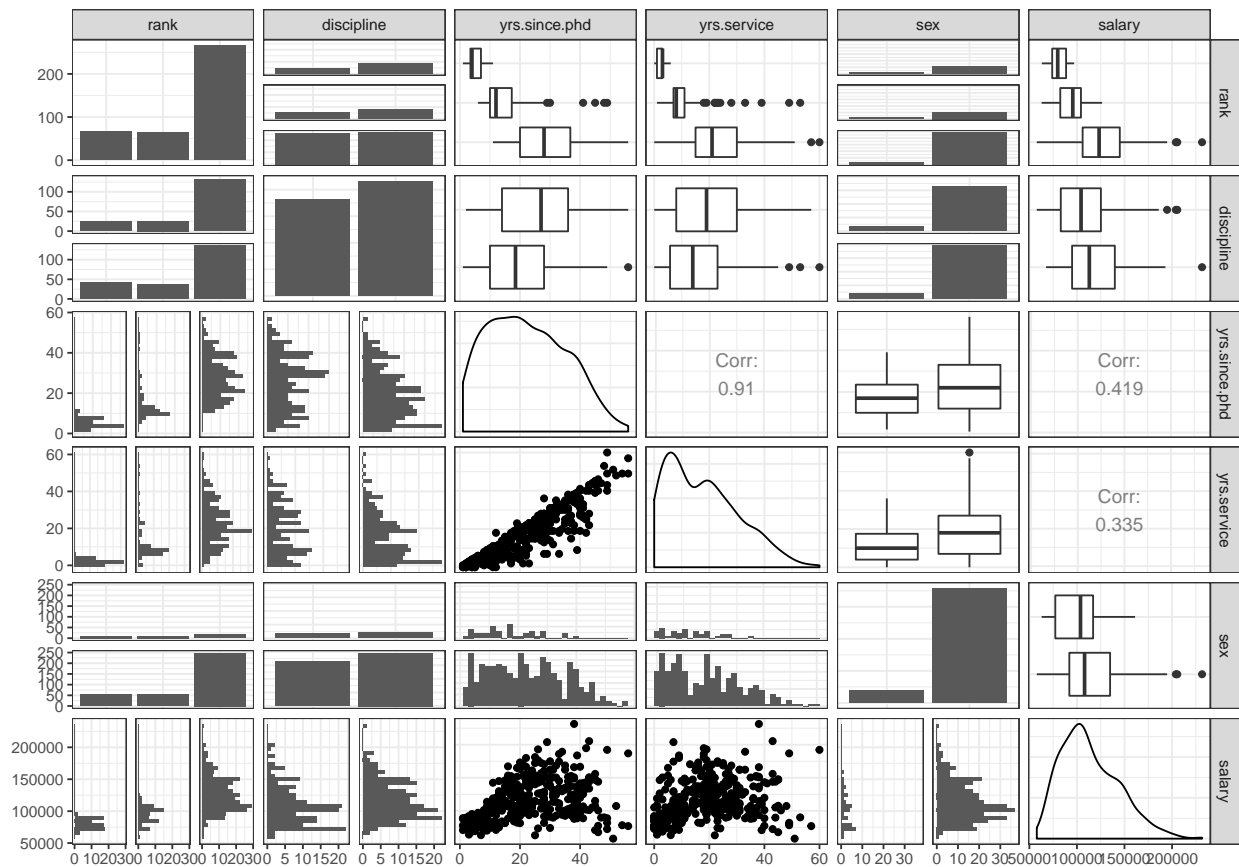
Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 91718.7 2765.8 33.162 <2e-16 *yrs.since.phd* 985.3 107.4 9.177 <2e-16 — Signif. codes: 0 ‘**0.001**’ ‘**0.01**’ ‘**0.05**’ ‘**0.1**’ ‘**1**’

Residual standard error: 27530 on 395 degrees of freedom Multiple R-squared: 0.1758, Adjusted R-squared: 0.1737 F-statistic: 84.23 on 1 and 395 DF, p-value: < 2.2e-16

### Question 12:

The following diagram shows all bivariate relationships.



### Question 13:

The next table shows mean salaries by rank and sex.

##	sex	
## rank	Female	Male
## AsstProf	78049.91	81311.46
## AssocProf	88512.80	94869.70
## Prof	121967.61	127120.82

#### Question 14:



#### Question 15:

We used a Chi-square test because it is appropriate for testing differences in proportions. We can not reject the null hypothesis of the Chi-square test since the p-value of the test is larger than 0.05. This means that we can not conclude that there is a difference between the proportions of women in each discipline.

```
##
##           A   B
## Female   18  21
## Male    163 195

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  cont
## X-squared = 2.0875e-29, df = 1, p-value = 1
```