SWP1 - Group 6

Thomas Siskos, Julian Cantor, Oliver Brose, Konstantin Haeusler

March 3, 2018

```
library(car)
library(dplyr)
library(ggplot2)
library(GGally)
data("Salaries")
```

1:

Observations: 397 - Variables: 6

```
dim(Salaries)
```

```
## [1] 397   6
```

2:

Nr. of professors with more than 40 yrs of experience: 21

```
old = Salaries[, "yrs.service"] > 40
sum(old)
```

```
## [1] 21
```

3:

Nr. of professors with salary higher than 150000: 54

```
rich = Salaries[, "salary"] > 150000
sum(rich)
```

```
## [1] 54
```

4:

Mean salary of professors with more than 20 yrs of experience: $122103.90

```
mean(Salaries[Salaries$yrs.service > 20, "salary"])
```

```
## [1] 122103.9
```

5:

```
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
```

**6:**

```
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
```
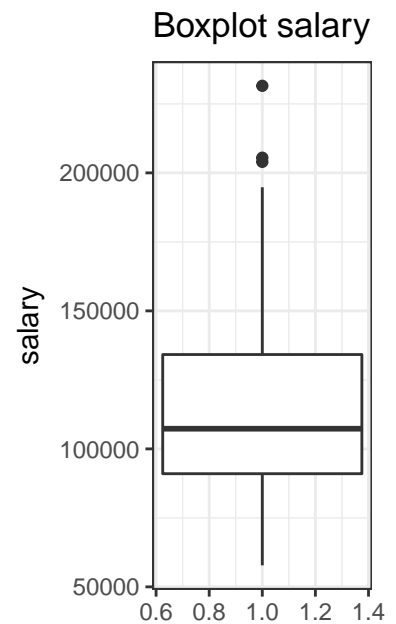
7:

```
ggplot(Salaries, aes(x = yrs.service)) +
  geom_histogram(aes(y = ..density..), binwidth = 5) +
  geom_density(color="red") +
  theme_bw() +
  ggtitle("Distribution of years of service")
```
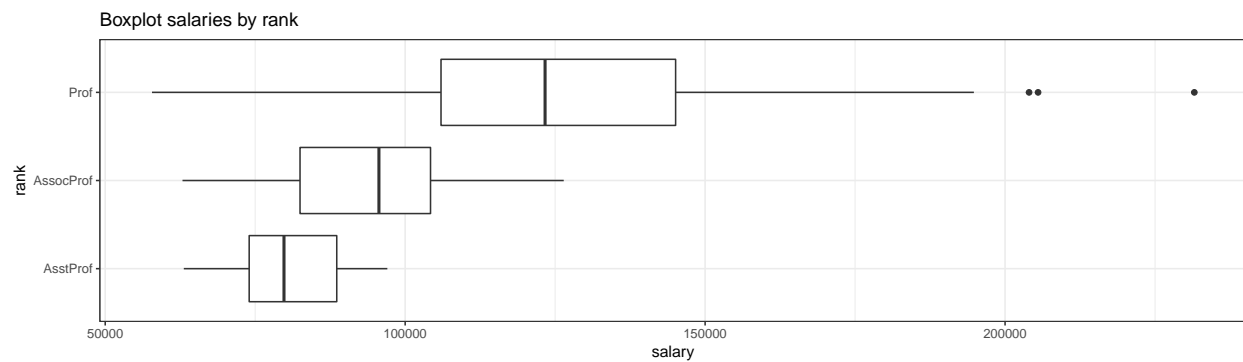
Distribution of years of service
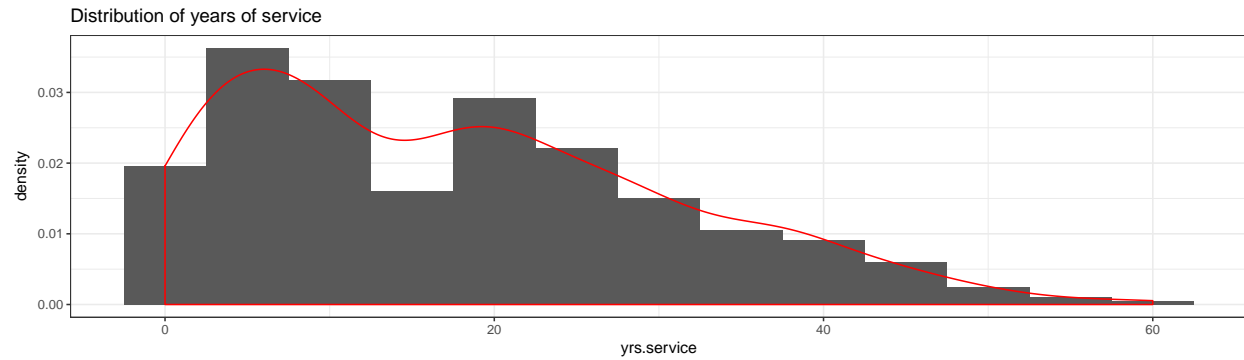
8:

```
ggplot(Salaries, aes(x = 1, y = salary)) +
  geom_boxplot() +
  theme_bw() +
  ggtitle("Boxplot salary") +
  xlab("")
```
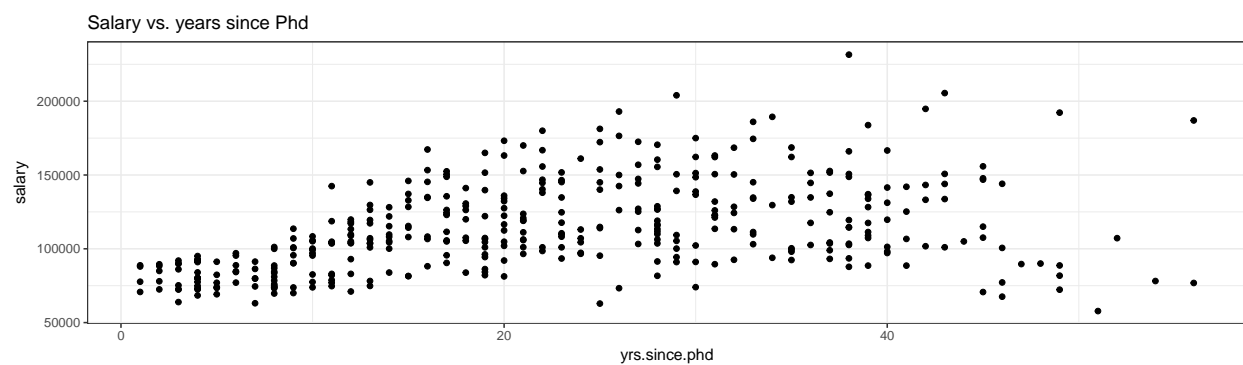

Boxplot salary

9:

```
ggplot(Salaries, aes(x = rank, y = salary, group = rank)) +
  geom_boxplot() +
  coord_flip() +
  theme_bw() +
  ggtitle("Boxplot salaries by rank")
```

Boxplot salaries by rank

10:

```
ggplot(Salaries, aes(x = yrs.since.phd, y = salary)) +
  geom_point() +
  theme_bw() +
  ggtitle("Salary vs. years since Phd")
```


Salary vs. years since Phd

11:

Correlation between salary and years since phd: 0.41

```
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)
summary(model)
```

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

4

```
## 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
```

12:

```
ggpairs(Salaries) +
  theme_bw()
```
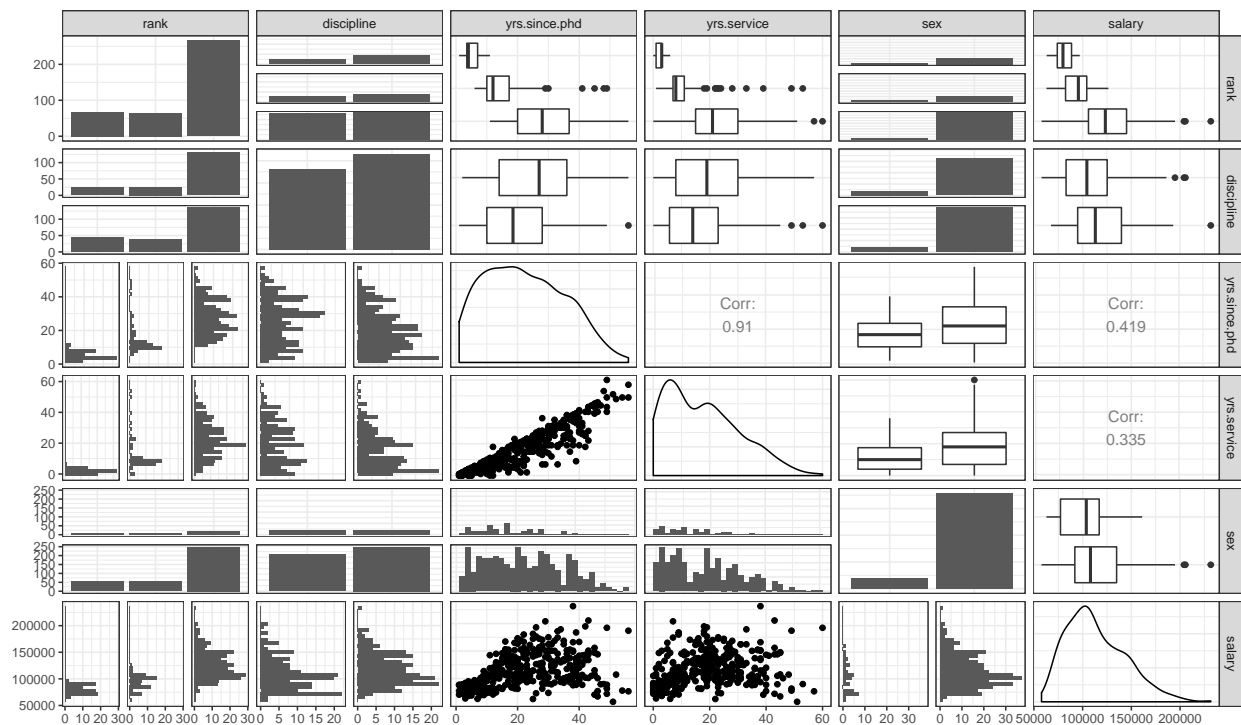

13:

```
Salaries %>%
  group_by(rank, sex) %>%
  summarise(mean.salary = mean(salary))
```

```
## # A tibble: 6 x 3
## # Groups:   rank [?]
##   rank     sex    mean.salary
##   <fct>    <fct>        <dbl>
## 1 AsstProf Female      78050.
## 2 AsstProf Male        81311.
## 3 AssocProf Female     88513.
## 4 AssocProf Male       94870.
## 5 Prof     Female     121968.
## 6 Prof     Male       127121.
```

14:

```r
ggplot(Salaries, aes(x = sex, y = salary, group = sex)) +
  geom_boxplot() +
  theme_bw() +
  coord_flip() +
  ggtitle("Distribution of salaries conditioned on sex")
```


15:
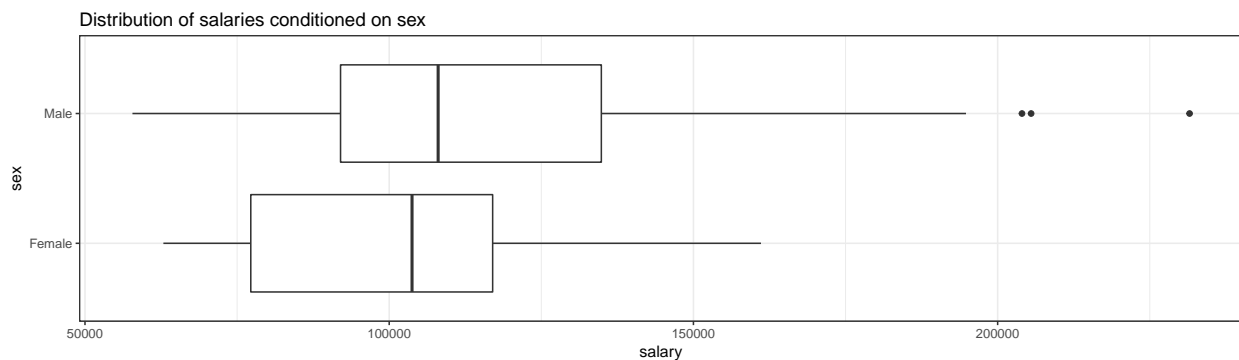
```r
cont = table(Salaries$sex, Salaries$discipline)
chisq.test(cont)
```

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