STAT 301 Worksheet 1

Peyton Hall

2025-01-17

```
finalscores <- c(90,78,92,69,85,88,83,96,72,99)
finalscores
```

```
##  [1] 90 78 92 69 85 88 83 96 72 99
```

```
mean(finalscores)
```

```
## [1] 85.2
```

```
median(finalscores)
```

```
## [1] 86.5
```

```
sd(finalscores)
```

```
## [1] 9.874771
```

```
min(finalscores)
```

```
## [1] 69
```

```
max(finalscores)
```

```
## [1] 99
```

```
sum(finalscores)
```

```
## [1] 852
```

```
# display a summary of the finalscores vector
summary(finalscores)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   69.00   79.25   86.50   85.20   91.50   99.00
```

```r
# create a data frame with three variables

Name <- c("John", "Kate", "Tom", "Cindy", "Jack", "Hana")
title <- c("Manager", "Prof", "Prof", "Manager", "Prof", "Manager")
Salary <- c(70000, 55000, 60000, 80000, 50000, 90000)

salarydf <- data.frame(Name, title, Salary)
salarydf
```

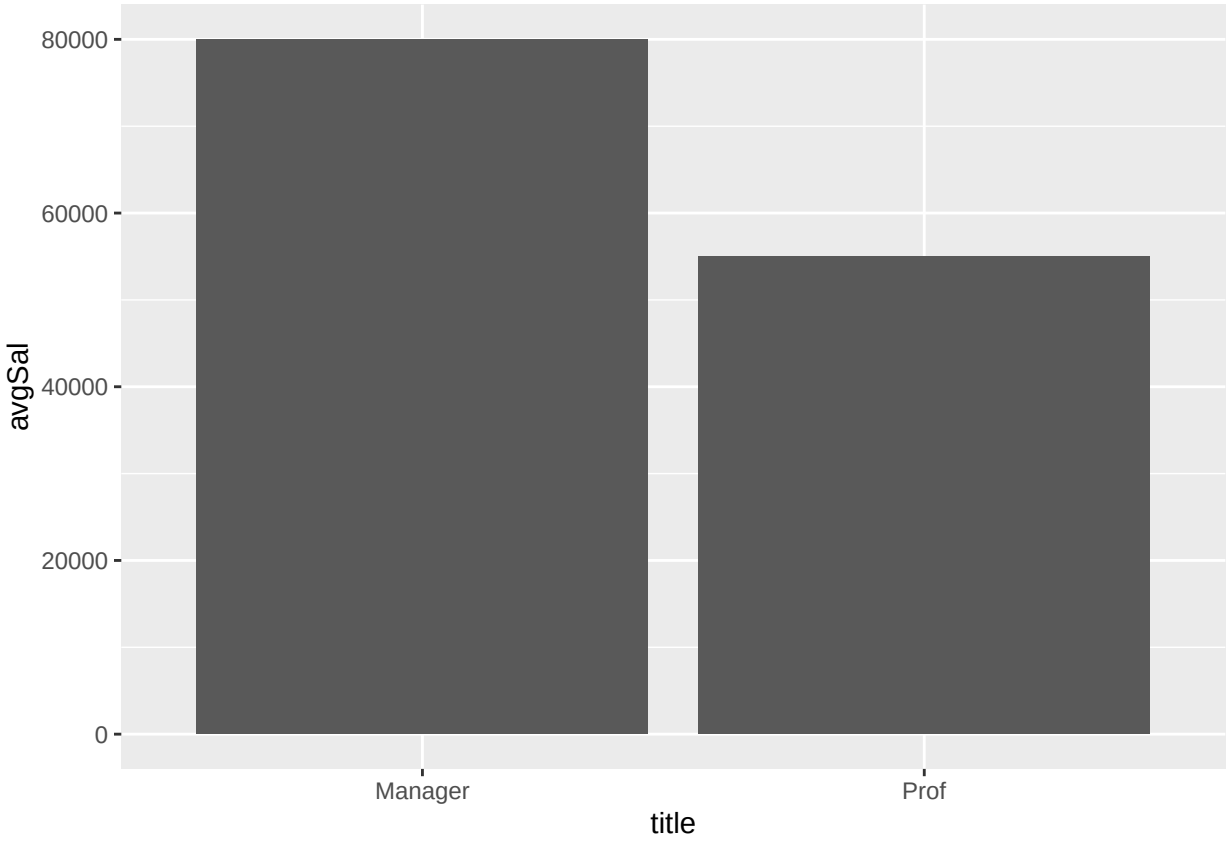
```
##    Name   title Salary
## 1  John Manager  70000
## 2  Kate    Prof  55000
## 3   Tom    Prof  60000
## 4 Cindy Manager  80000
## 5  Jack    Prof  50000
## 6  Hana Manager  90000
```

```r
# filter the data frame for managers only
managerdf <- salarydf %>%
  filter(title == "Manager")

managerdf
```

```
##    Name   title Salary
## 1  John Manager  70000
## 2 Cindy Manager  80000
## 3  Hana Manager  90000
```

```r
# generate a bar graph of average salaries by title
salarydf %>%
  group_by(title) %>%
  summarize(avgSal = mean(Salary)) %>%
  ggplot(aes(x=title, y=avgSal)) + geom_bar(stat = "identity")
```




```
# install.packages("tidyverse")
library(tidyverse)

# select only title and Salary columns from salarydf
newsalarydf <- salarydf %>%
  select(title, Salary)

newsalarydf
```

```
##     title Salary
## 1 Manager  70000
## 2    Prof  55000
## 3    Prof  60000
## 4 Manager  80000
## 5    Prof  50000
## 6 Manager  90000
```

```
# create a data frame for the activity
# a) Create data frame
Name1 <- c("John", "Peter", "Jolie", "Jason", "Leslie", "Donna")
gender <- c("Male", "Male", "Female", "Male", "Female", "Female")
Salary1 <- c(21000, 23000, 25000, 30000, 20000, 35000)
Starting_Date <- c("2010-11-01", "2008-12-15", "2007-12-01", "2001-10-10", "2009-01-10", "2001-05-01")

chef <- data.frame(Name1, gender, Salary1, Starting_Date)
chef
```

```
##    Name1 gender Salary1 Starting_Date
## 1   John   Male   21000    2010-11-01
## 2  Peter   Male   23000    2008-12-15
## 3  Jolie Female   25000    2007-12-01
## 4  Jason   Male   30000    2001-10-10
## 5 Leslie Female   20000    2009-01-10
## 6  Donna Female   35000    2001-05-01
```

```
# b) Use chef data
newchef <- chef %>%
  select(gender, Salary1, Starting_Date)
newchef
```

```
##   gender Salary1 Starting_Date
## 1   Male   21000    2010-11-01
## 2   Male   23000    2008-12-15
## 3 Female   25000    2007-12-01
## 4   Male   30000    2001-10-10
## 5 Female   20000    2009-01-10
## 6 Female   35000    2001-05-01
```

```
# c) Keep females only
femalesdf <- chef %>%
  filter(gender == "Female")
femalesdf
```

```
##    Name1 gender Salary1 Starting_Date
## 1  Jolie Female   25000    2007-12-01
## 2 Leslie Female   20000    2009-01-10
## 3  Donna Female   35000    2001-05-01
```

```
# d) Create new column
chef %>%
  group_by(gender) %>%
  summarize(avgSal = mean(Salary1), sdSalary = sd(Salary1))
```

```
## # A tibble: 2 x 3
##   gender avgSal sdSalary
##   <chr>   <dbl>    <dbl>
## 1 Female 26667.    7638.
## 2 Male   24667.    4726.
```