R Notebook

```
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --

## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.3      v dplyr   1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Get data and preliminary prep

```r
# setwd(choose.dir())
d <- read_csv("MKT 382 data 2021 Survey 2.csv")

## Rows: 54 Columns: 20

colnames(d) <- c(
  "timestamp",
  "computer",
  "phone",
  "sports",
  "degree",
  "humor",
  "R",
  "SQL",
  "Python",
  "programming",
  "queso",
  "band",
  "instructor",
  "home",
  "city",
  "math",
  "sleep",
  "late",
  "vegetarian",
  "siblings"
)
```

Problems and solutions

Problem 1

Generate a data frame (tibble or equivalent) showing the top three `siblings` values (in descending order). The data frame should have only one column: `siblings`. (You can ignore ties.)

```
## # A tibble: 3 x 1
##    siblings
##       <dbl>
## 1         6
## 2         4
## 3         4
```

I'm scared of ties.

Problem 2

Generate a data frame showing the number of survey responses by degree in descending order of those counts.

```
## # A tibble: 7 x 2
##    degree                               count
##    <chr>                                <int>
## 1 Business                                22
## 2 Advertising, communications             19
## 3 Science or math                          6
## 4 Liberal arts (languages, history, etc)   4
## 5 Economics, policy, government            1
## 6 Engineering                              1
## 7 <NA>                                     1
```

Problem 3

Generate a data frame of the mean instructor greatness responses by degree. The data frame should be in descending order of mean greatness.

```
## # A tibble: 7 x 2
##    degree                               mean_greatness
##    <chr>                                         <dbl>
## 1 Economics, policy, government                     5
## 2 Engineering                                       5
## 3 Liberal arts (languages, history, etc)            5
## 4 Science or math                                   5
## 5 <NA>                                              5
## 6 Advertising, communications                    4.95
## 7 Business                                       4.86
```

Problem 4

For Business majors, the count of responses by computer type.

```
## # A tibble: 2 x 2
##   computer    count
##   <chr>       <int>
## 1 Mac            11
## 2 Windows PC     11
```

Problem 5

For Business majors with Macs, the mean number of siblings by queso.

```
## # A tibble: 2 x 2
##   queso         mean_siblings
##   <chr>                 <dbl>
## 1 Kirby Lane's              2
## 2 Torchy's                2.6
```
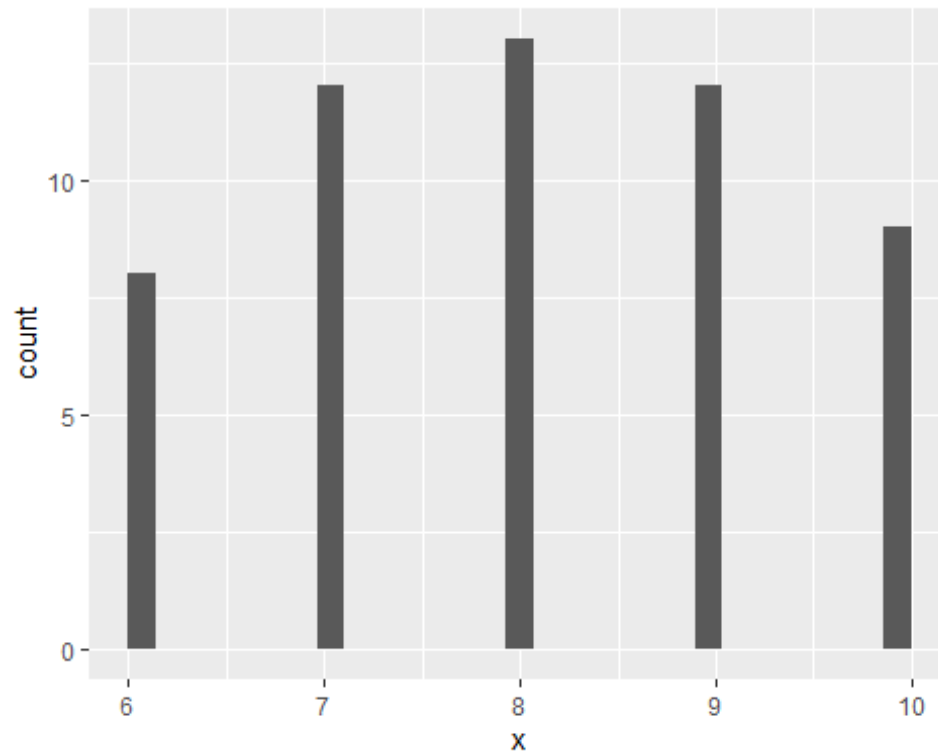
Problem 6

Consider a strange metric that is math+instructor. This metric, which of course is the sum of the "how much you like math" response and the "how great is the instructor response", might have tremendous predictive value. Let's investigate.

Compute the mean value of this metric by computer value.

```
## # A tibble: 3 x 2
##   computer         mean_x
##   <chr>             <dbl>
## 1 Both PC and Mac    8
## 2 Mac                8.09
## 3 Windows PC         7.94
```

Problem 7

Generate a ggplot histogram of the math+instructor values.

Problem 8

For responses reporting one or two siblings, count votes for each queso.

```
## # A tibble: 3 x 2
##   queso            n
##   <chr>        <int>
## 1 Kirby Lane's     8
## 2 Torchy's        31
## 3 <NA>             1
```