Homework 1

1. Compute
$$1^2 + 2^2 + 3^2 + \cdots + 99^2 + 100^2.$$

```
vec <- 1:100
sum(vec^2)
```

```
## [1] 338350
```

2. How many digits does $N = 1234567891011121314...9998999910000$ have, where N is formed by concatenating $1, 2, 3, \ldots, 9999, 10000$? (*Hint*: you might want to look up the functions `nchar()`, `as.character()`, `paste()`)

```
vec_char <- as.character(1:10000) # turn numeric to character
vec_collapse <- paste(vec_char, collapse = "")
print(nchar(vec_collapse))
```

```
## [1] 38894
```

`iris` is a data set that comes with base R. It consists of 150 observations with 5 columns, Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species, which are features related to Iris flowers. You can view the first few rows using `head()`:

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

3. Compute the sample mean of `Sepal.Length`. Also, what is the *fifth* smallest value of `Sepal.Length`?

```
mean(iris$Sepal.Length)
```

```
## [1] 5.843333
```

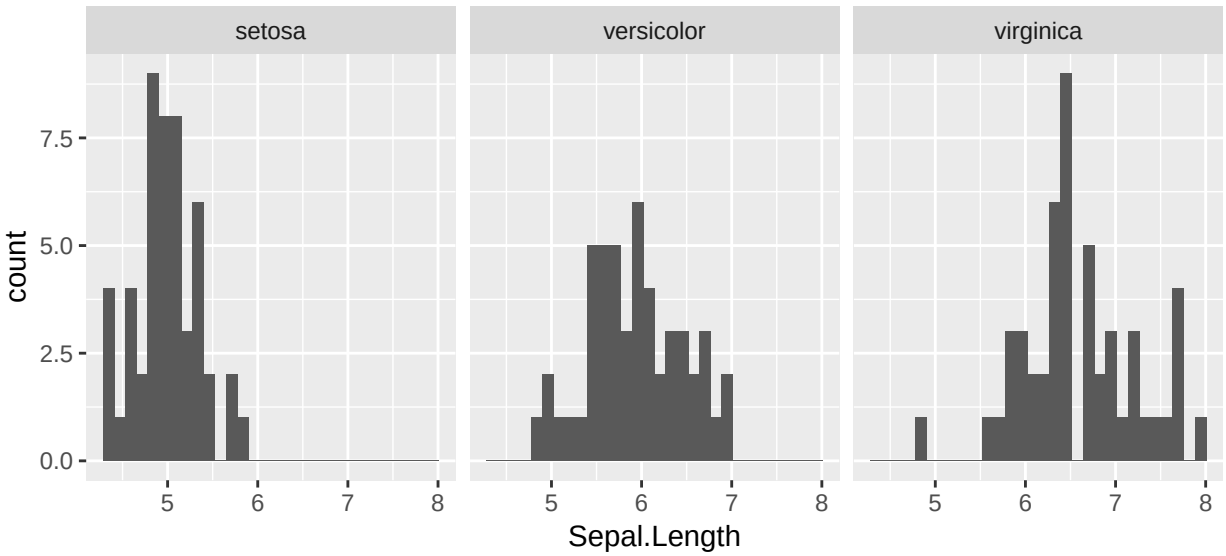
```
sort(iris$Sepal.Length)[5]
```

```
## [1] 4.5
```

4. Obtain the following plots that describe the histogram of `Sepal.Length` by `Species`.

```
g <- ggplot(iris, aes(Sepal.Length)) +
  geom_histogram() +
  facet_wrap(~Species)

print(g)
```



5. Recall that the data set `heights.txt` consists of 1375 observations of mother/daughter height pairs. As done in class, you can read in the data set using the following code:

In how many pairs is mother's height greater than daughter's height?

```
sum(d$Mheight > d$Dheight)
```

```
## [1] 399
```