# Homework 3

due Feb 2, 2021

(Problems **1**, **2**) Recall `iris` data set.

1. Find the maximum value of the sum `Sepal.Length` + `Sepal.Width`. To which species does the sample with the maximum value you found above belong? *Hint*: you might want to start with defining a new variable `Sepal.Sum` using `mutate()`, for example.

2. Create the following plot. Note that `setosa` species do *not* show up. You do not need to adjust size or width/height ratio of the plot.
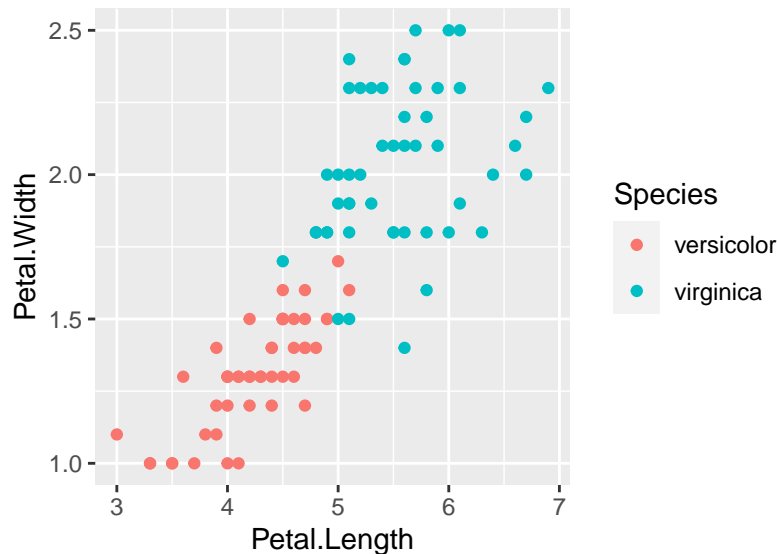
(Problems **3**, **4**, **5**) Recall the `heights.txt` data we used in class with 1375 mother/daughter height pairs. The following code reads in the data set as `ht`.

```
ht <- read.table(file = "heights.txt", header = T, sep = " ")
```

3. Define an extra variable `FamilyID` for each family, like `Fam_1`, `Fam_2`, ..., `Fam_1375`, so that `head(ht)` and `tail(ht)` return the following, respectively.

```
##    Mheight Dheight FamilyID
## 1     59.7    55.1    Fam_1
## 2     58.2    56.5    Fam_2
## 3     60.6    56.0    Fam_3
## 4     60.7    56.8    Fam_4
## 5     61.8    56.0    Fam_5
## 6     55.5    57.9    Fam_6
```

```
##      Mheight Dheight FamilyID
## 1370   69.5    70.4 Fam_1370
## 1371   69.1    70.1 Fam_1371
## 1372   65.0    71.6 Fam_1372
## 1373   66.3    71.4 Fam_1373
## 1374   70.8    71.0 Fam_1374
## 1375   63.0    73.1 Fam_1375
```

4. In how many families, is mother taller than daughter?

5. Create the following box plot. *Hint*: you might want to start with transforming `ht` into a long form.