

# Quiz 1: STAT 3622 Data Visualization

*19 September 2016 (Monday)*

## 1. Stem-and-Leaf Plot (3pt)

The stem-and-leaf plot for a numerical variable splits the scores into **stem** (the first digit/digits, listed downward) and **leaf** (usually the last digit, listed rightward). For example, we may take a random sample of `iris$Sepal.Length` and use the R function `stem` to display the stem-and-leaf plot as follows.

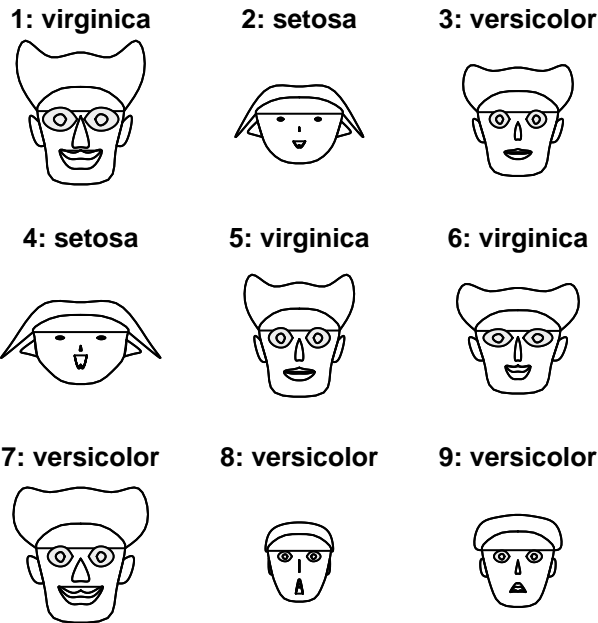
```
##
## The decimal point is at the |
##
## 4 | 3
## 4 | 6899
## 5 | 11
## 5 | 578889
## 6 | 13
## 6 | 5677889
## 7 | 122
## 7 | 77
```

Convert it to the histogram with 4 breaks (i.e. bins). Draw the picture.

## 2. Chernoff Faces (4pt)

Chernoff faces, invented by Herman Chernoff (currently at Harvard), may display multivariate data in terms of a human face, including shape, size, placement and orientation of face, eyes, ears, mouth, nose and hair, as well as emotional expression. One may compare the samples based on the changes/differences among such *statistical* faces.

In R, we may use the `faces` function from the `aplpack` package to plot Chernoff faces, e.g. using a random sample of 9 observations of `iris` dataset.



Rank order these 9 samples based on different parts of faces. For example: height of nose  $1 > \{5,6\} > 7 > 3 > \{9,8\} > \{2,4\}$  (this feature can be skipped in your answer.)

### 3. Data Science Workflow (3pt)

Suppose you are a data scientist at Facebook responsible for an image recognition and tagging project. You need to design intelligent algorithms for 1) recognizing faces and 2) suggesting tags, when a new image is uploaded; see an illustration below. How would you approach to solving this problem in different steps?

