STA 326 2.0 Programming and Data Analysis with R *

Exploring `iris` dataset with `qplot`

30 March 2020

Stage 1: Planning your analysis

Step 1: Dataset overview and description

Before we get started let's look at the data and plan the analysis.

Load iris dataset

```
data(iris)
```

Here is a glimpse of the dataset

```
head(iris)
```

```
  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
```

We have four qunatitative variables: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width and one qualitative variable Species

Step 2: Let's look at the graphs we could use to explore variables one by one.

Plots that could be used to to summarize quantitative variables

- Box and whisker plot
- Histograms
- Dotplots
- Density plot
- Stem and leaf displays

Note: Stem and leaf displays are best-suited for small to moderate datasets, whereas others such as histograms and Box and whisker plots are best-suited for large datasets. Graph types such as Box and whisker plots are good at depicting differences between distributions and identyfying outliers.

Plotss that could be used to to summarize qualitative variables

- Bar chart
- Pie chart

Step 3: Next we will look at two variables at a time.

- Quantitative and Quantitative: Scatter plots
- Quantitative and Qualitative: Boxplots/ Histograms/ Dotplots/ Density plots with groups allow us to compare across different levels of the qualitative variable. **Faceting** can be used to generate the same plot for different levels of the qualitative variable.

Step 4: Three variables at a time.

- Two quantitative variables and one qualitative variable: Scatterplot with different markers (eg: size, shapes, colours) for different levels of the qualitative variable.

Stage 2: Getting started with `qplot()` in the ggplot2 package.

Now we are going to use the `qplot` function to make some quick plots. In your final analysis you do not need to include all the graphs we discussed in Stage 1. However, here I plot different graphs under same category . It helps demonstrate how different graphs can be plotted for various purposes using the `qplot`.

Recap: qplot syntax

```
qplot(
  x,
  y,
  data,
  facets = NULL,
  margins = FALSE,
  geom = "auto",
  xlim = c(NA, NA),
  ylim = c(NA, NA),
  log = "",
  main = NULL,
  xlab = NULL,
  ylab = NULL,
  asp = NA,
  stat = NULL,
  position = NULL
)
```