Created by Sean Leggett, January 29, 2020

BDA201 2020 Winter Session – Weekly Assignment 1

Question 1
What type of plot is this?

Answer:
The plot provided in the example is a line plot showing two variables: mobile phone antennae per 1,000 people and kidnapping rates per 100,000 people. Both for the country of Colombia. The line plot is a time series.

Question 2
Describe the phenomenon that is shown in the plot.

Answer:
By using two variables on the plot, it implies a relationship between increasing cell phone availability and reduction of kidnapping rates. However, demonstrating correlation is not evidence of causation. Possibly more phones means kidnappers are hesitant to act in public. More likely perhaps is 3rd variable coinciding such as increased police presence on the streets or increasing income levels. Perhaps even a combination of the above.

Question 3
Which plot types asks you to draw a cause and effect relationship?

Answer:
A scatter plot is ideal for plotting relationship between two variables. It demonstrates correlation but not necessarily causation.
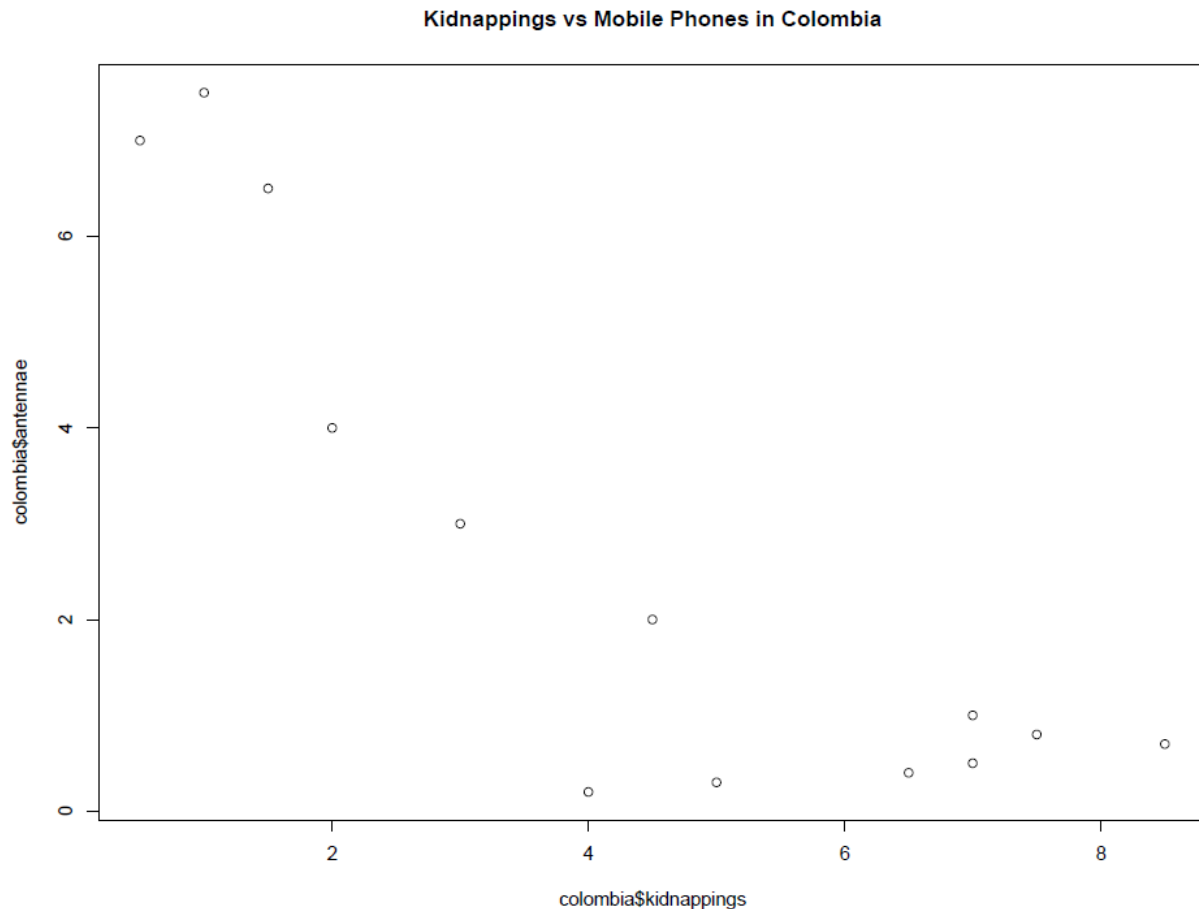
Question 4
Provide an estimated version of the plot proposed.

Answer:
Below is code to produce a scatter plot of the same (estimated) data in the example plot.

```
colombia <- data.frame(
  year = c("1996", "1997", "1998", "1999", "2000", "2001", "2002", "2003", "2004", "2005", "2006",
"2007", "2008"),
  kidnappings = c(4, 5, 6.5, 7, 8.5, 7.5, 7, 4.5, 3, 2, 1.5, 1, 0.5),
  antennae = c(0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 1, 2, 3, 4, 6.5, 7.5, 7)
)
colombiaplot <- plot(colombia$kidnappings, colombia$antennae,
          main = "Kidnappings vs Mobile Phones in Colombia")
```

Output for the code above is as follows:

Kidnappings vs Mobile Phones in Colombia

Question 5
What advantages does the provided plot have vs our plot?

Answer:
The original plot displays the same trend as the scatter plot (decreased kidnappings coinciding with increased mobile phone availability). However, by combining the two variables with the time series on X axis, the trend is certainly more obvious in a visual sense.

Question 6
Provide plot showing variability in web traffic for each day of the week.

Answer:
```
traffic <- read.csv("website-traffic.csv")
traffic$DayOfWeek = factor(traffic$DayOfWeek, c("Monday","Tuesday","Wednesday", "Thursday",
"Friday", "Saturday", "Sunday"))
trafficplot <- boxplot(Visits~DayOfWeek, data = traffic,
             main = "Site Visit Variability by Day of Week")
```

NB: factor order for DayOfWeek reset to show week starting Monday in boxplot

Output for this code is shown below:

Site Visit Variability by Day of Week
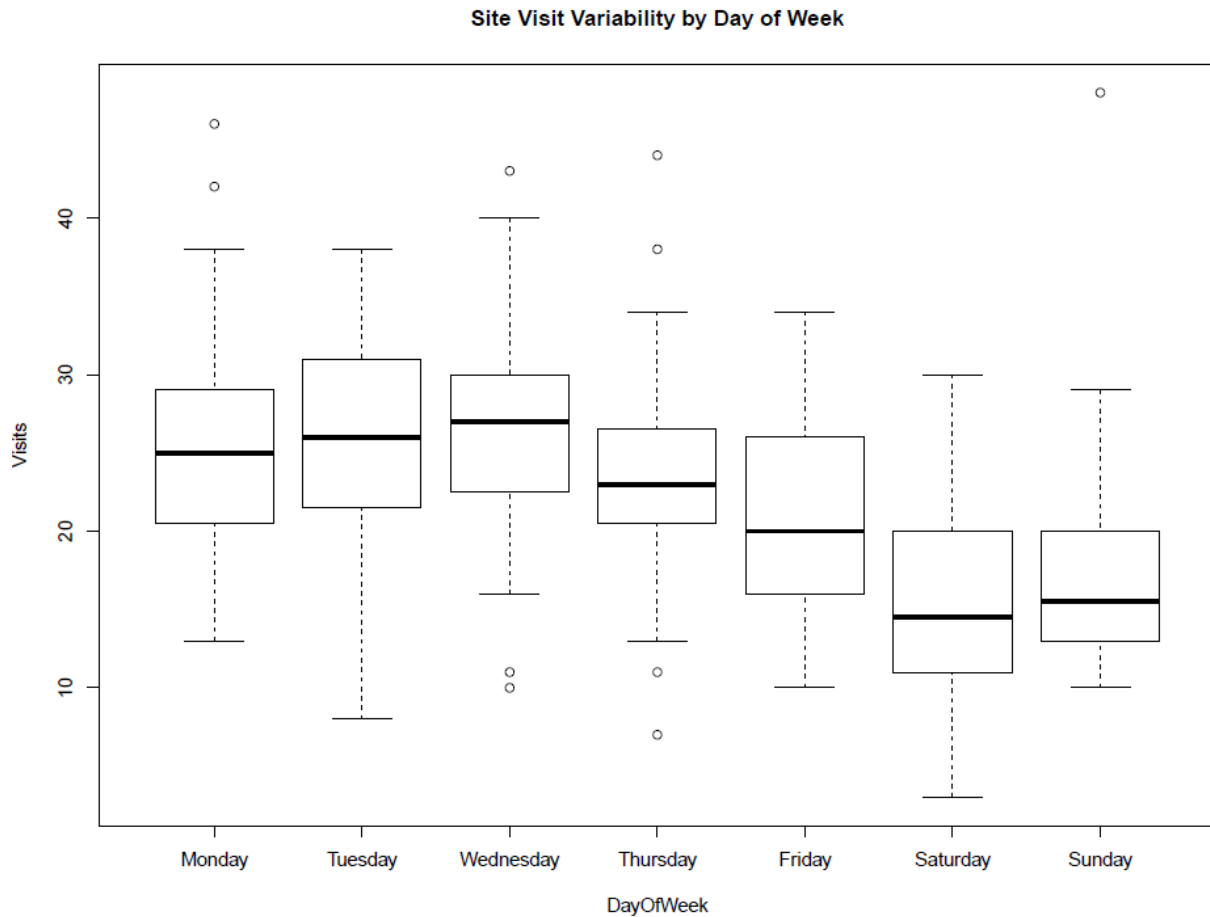
Reference Material:

1. Course Content
2. <u>Sams Teach Yourself R in 24 Hours</u>, Andy Nicholls, Richard Pugh, Aimee Gott. Sams, 2016.
3. https://stackoverflow.com/questions/4260698/r-ordering-in-boxplot