BDA201 - Assignment 1

Created by Sean Leggett, January 29, 2020 BDA201 2020 Winter Session

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 is implying a relationship between the two in a visual sense only. Rates may well be going down for kidnappings but relationship to cell phone availablility may be tenuous at best. The rate of change may be relevant but the hard scales of per 100,00 vs per 1,000 are substantially different.

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

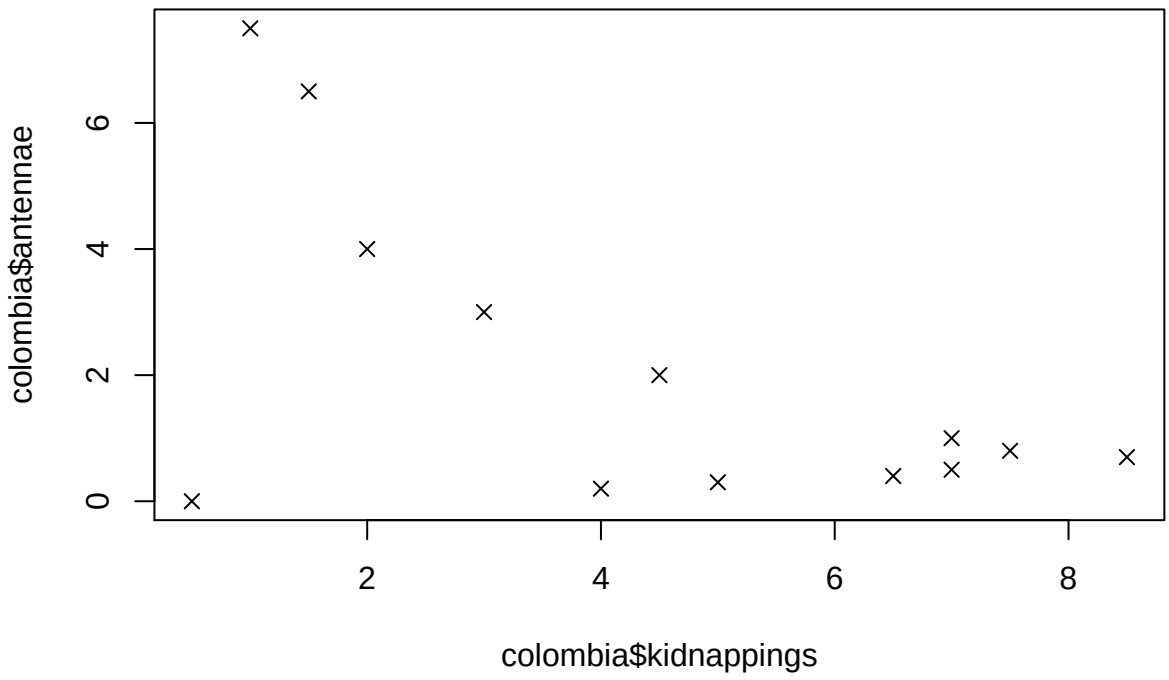
A scatter plot is ideal for plotting relationship between two variables.

Question 4 Provide an estimated version of the plot proposed.

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", "200
  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, 0)
)

colombiaplot <- plot(colombia$kidnappings, colombia$antennae, pch = 4)
```




```
colombiaplot
```

```
## NULL
```

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

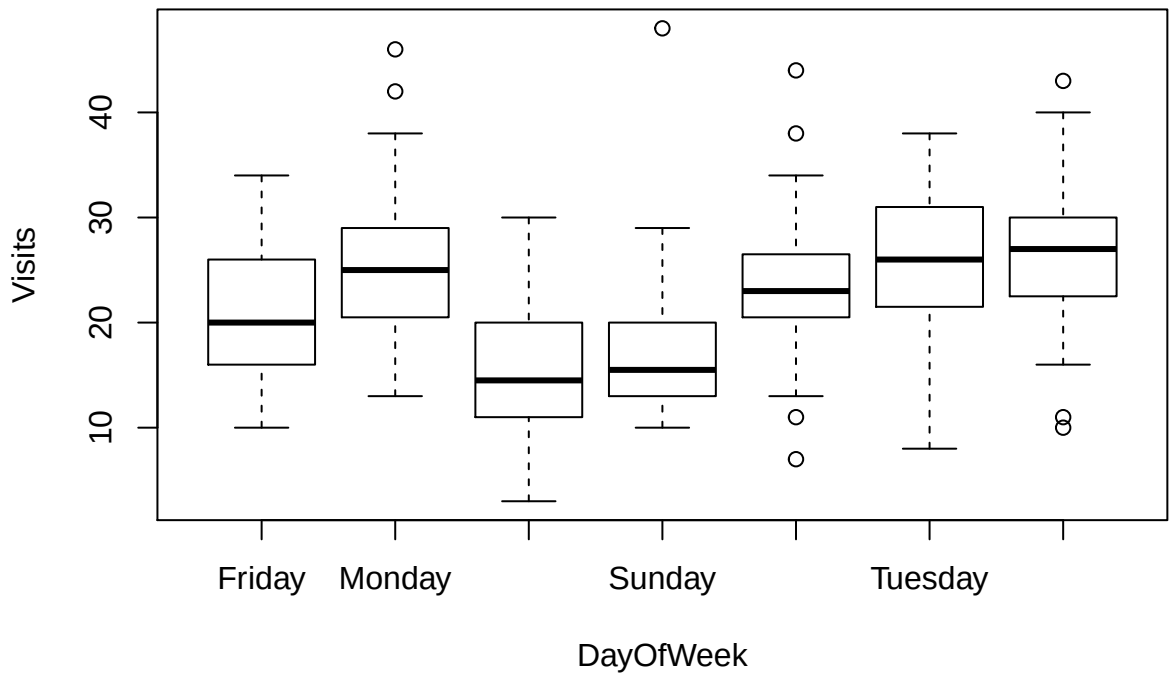
Answer: The original plot displays the same trend as the scatter plot (decreased kidnappings coinciding with increased mobile phone availability). However, by combining the two variables with the time series on X axis, the trend is certainly more obvious to the untrained eye.

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

Answer:

```
traffic <- read.csv("website-traffic.csv")

trafficplot <- boxplot(Visits~DayOfWeek, data = traffic)
```




```
trafficplot
```

```
## $stats
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]   10 13.0  3.0 10.0 13.0  8.0 16.0
## [2,]   16 20.5 11.0 13.0 20.5 21.5 22.5
## [3,]   20 25.0 14.5 15.5 23.0 26.0 27.0
## [4,]   26 29.0 20.0 20.0 26.5 31.0 30.0
## [5,]   34 38.0 30.0 29.0 34.0 38.0 40.0
## attr(,"class")
##    Friday
## "integer"
##
## $n
## [1] 30 31 30 30 31 31 31
##
## $conf
##          [,1]    [,2]    [,3]     [,4]     [,5]     [,6]     [,7]
## [1,] 17.11533 22.5879 11.9038 13.48073 21.29734 23.30412 24.87168
## [2,] 22.88467 27.4121 17.0962 17.51927 24.70266 28.69588 29.12832
##
## $out
##  [1] 46 42 48 38 11 44  7 11 43 10
##
## $group
##  [1] 2 2 4 5 5 5 5 7 7 7
```

```
##
## $names
## [1] "Friday"    "Monday"    "Saturday"  "Sunday"    "Thursday"  "Tuesday"
## [7] "Wednesday"
```