Advanced graphics in R using ggplot2

Stephen A. Sefick

2016-11-16

1) Using ggplot2

Background

In this exercise, you will explore a dataset that is included with R by making a variety of graphs using the ggplot2 package inside of R studio. Specificaly, we will investigate how a number of variables affect miles per gallon (mpg).

Explaination of the mtcars data: The data was extracted from the 1974 **Motor Trend** US magazine,and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models) (From: ?mtcars).

Exercise

Explore the use of the ggplot2 package.

1. Please create a file named **your_last_name_Ex1.R**, and write all of the code used for this exercise there. In addition, you will be outputing a selection of the graphs you developed for this exercise into a single pdf file (see 18 below).
2. Install ggplot2 and grid.

   ```
   install.packages(c("ggplot2", "grid"))
   ```

   - The online documentation for ggplot2 is: ggplot2
3. Load the ggplot2 package and the mtcars data.

   ```
   library(ggplot2)
   data(mtcars)
   ```
4. Look at the transmission type column (am). Now, change transmission type to a more intuitive coding.

```
mtcars[mtcars$am==1,"am"] <- "automatic"
mtcars[mtcars$am==0,"am"] <- "manual"
```

5. In order to access how the data might be distributed, make a histogram of mpg, and color it by transmission type.

```
qplot(mpg, data=mtcars, geom="histogram", bins=5, fill=am, col=am)
```

6. Make 2 boxplots to explore the effect of number of gears and cylendars on mpg. Use the **geom** argument. Use the theming system to change the default look of ggplot2. Because the number of cylendars and gears are numeric (e.g., str(mtcars$cyl)), make cyl and gear into a factor (i.e., qplot(as.factor(cyl), mpg, data=mtcars, geom="boxplot")).

```
#Theming system example. See the documentation for more information
qplot(as.factor(c(rep("A",5), rep("B", 5))), 1:10, geom="boxplot")+theme_bw()
```

   - **Notice the x axis label is non-sense.** It is probably a good idea to fix this. Use xlab in the qplot command to fix this (look at the documentation of qplot if you are having problems).

7. We might be interested in whether automatic and manual transmission have the same relationships. You want to use the facet_wrap() functionality.

```
#Theming system example. See the documentation for more information
your_plot+facet_wrap(~am)
```

8. Make a scatterplot to explore the relationship of horsepower and miles per gallon.

9. Use a smoother to explore this relationship. Specifically, use a lowess smoothing line.

```
#Theming system example. See the documentation for more information
your_plot+geom_smooth()
```

10. Do these data suggest a linear or some other type of relationship?

    - Try log transforming both mpg and hp

```
#log transform
qplot(log(hp), log(mpg), data=mtcars)+geom_smooth()
```

    - Is this relationship **linear now**?

11. We can use specific methods in geom_smooth(). Now we have a linear relationship after transformation. We can use a linear model to explore these data.

```
#log transform
qplot(log(hp), log(mpg), data=mtcars)+geom_smooth(method="lm")
```

12. Explore this relationship with transmission type included as a facet.

13. Does the intercept look the same?

14. Does the slope look the same?

15. The faceted graph is nice looking but it would be better to look at these relationships without the facet. This can be acomplished with setting color or shape. Color these graphs by transmission type.

```
#log transform
qplot(log(hp), log(mpg), data=mtcars, col=am)+geom_smooth()
```

16. Everything else being **equal**, and assuming you want to optimize **mpg**. What type of transmission would you choose?

17. Since this data is multivariate, use geom_grid() the relationship of hp with mpg but also adding the information of gear, transmission, and number of cylendars. With what we have learned previously and gridded faceting (i.e., facet_grid())

```
mpg_hp_gear_trans_cyl <- qplot(hp, mpg, data=mtcars, geom="blank")+facet_grid(gear~am)

mpg_hp_gear_trans_cyl+geom_smooth(method="lm")+geom_point(aes(color=as.factor(cyl)))
```

1. Please save your plots from 5, 6, 15, and 17 into a single pdf called **your_last_name_Ex1.pdf**.

```
#for example
pdf("your_last_name_Ex1.pdf")
plot_5
plot_6
plot_7
plot_15
plot_17
dev.off()
```

2) Time Series - Experiment Sandy Creek

Background

The data come from an experiment that I ran in 2013 in a creek near Waverly, AL. I used velocity (random slope) nested within block (random intercept) in a mixed model framework. I did this to statistically account for differences in velocity within block. This dataset has 3 columns.

Column Description

- **date** is date of velocity measurement
- **block** is experimental block

- **velocity** is mean velocity within a block (i.e., point measure of water movement in m/s)

Questions

1. Does velocity vary with time?
2. Does velocity vary among blocks?
3. There was a rain event during the experiment. When was this? (*HINT: uniform increase in velocity at all blocks*)

Exercise

Make a plot to investigate how velocity changes through time and block.

1. Create a file named **your_last_name_Ex2.R** in the **Time_Series** folder. Read in **experiment_velocity_time_series.csv** to an object called vel_exp

```
#must change date into class Date in order to have qplot recognize as such.
vel_exp <- read.csv("experiment_velocity_time_series.csv")
```

2. load the ggplot2 library, and change Date to Date class. This will ensure that ggplot2 can recognize this as a Date.

```
#must change date into class Date in order to have qplot recognize as such.
library(ggplot2)
vel_exp$date <- as.Date(vel_exp$date)
```

3. Plot the relationship of Date with velocity as a line graph using qplot (e.g., qplot(x, y, ...))
4. Use the col argument to color the lines by block
5. Decide if you want a facet or not (i.e., facet_wrap(~variable_of_interest)
6. Explore the "theme-ing" system theme_bw() or the custom theme that I have provided you.
7. To have the custom theme usable you will have to source the file **publication_ggplot2_theme.R** with the source command.

```
#use my publication theme
source("publication_ggplot2_theme.R")
your_plot+publication()
```

8. Once you have arrived at a suitable graph save it as a pdf called **your_last_name_Ex2.pdf**.

```
pdf("your_last_name_Ex2.pdf")
your_ggplot2_plot_object
dev.off
```

9. Please answer the questions posed above in a text file named **your_last_name_velocity_answers.txt**.

TURN IN OVER EMAIL

put your **your_last_name_Ex1.pdf**, **your_last_name_Ex2.pdf**, **associated R script files**, and **your_last_name_velocity_answers.txt** into a folder **named last_name**. gzip this file and email to sas0025@auburn.edu with the Subject: "**ggplot exercise - Dr. Stevision**".