

Lab Number 7

The R code below creates one of the dataframes (**dat**) that we will use for **Lab 7**. It is from chapter 4 of *Cookbook for R*. A link is here: [Cookbook for R](#)

```
library(tidyverse)
set.seed(955)
# Make some noisily increasing data
dat <- data.frame(cond = rep(c("A", "B"), each=10),
                  xvar = 1:20 + rnorm(20, sd=3),
                  yvar = 1:20 + rnorm(20, sd=3))

head(dat)
```

##	cond	xvar	yvar
## 1	A	-4.252354	3.473157275
## 2	A	1.702318	0.005939612
## 3	A	4.323054	-0.094252427
## 4	A	1.780628	2.072808278
## 5	A	11.537348	1.215440358
## 6	A	6.672130	3.608111411

Above, **rnorm** generates a vector of 20 normally distributed random numbers.

rnorm can take up to three arguments:

n: the number of random variables to generate (**20**).

mean, if not specified defaults to 0 (**not specified**).

sd: the standard deviation (**3**). If not specified, defaults to 1.

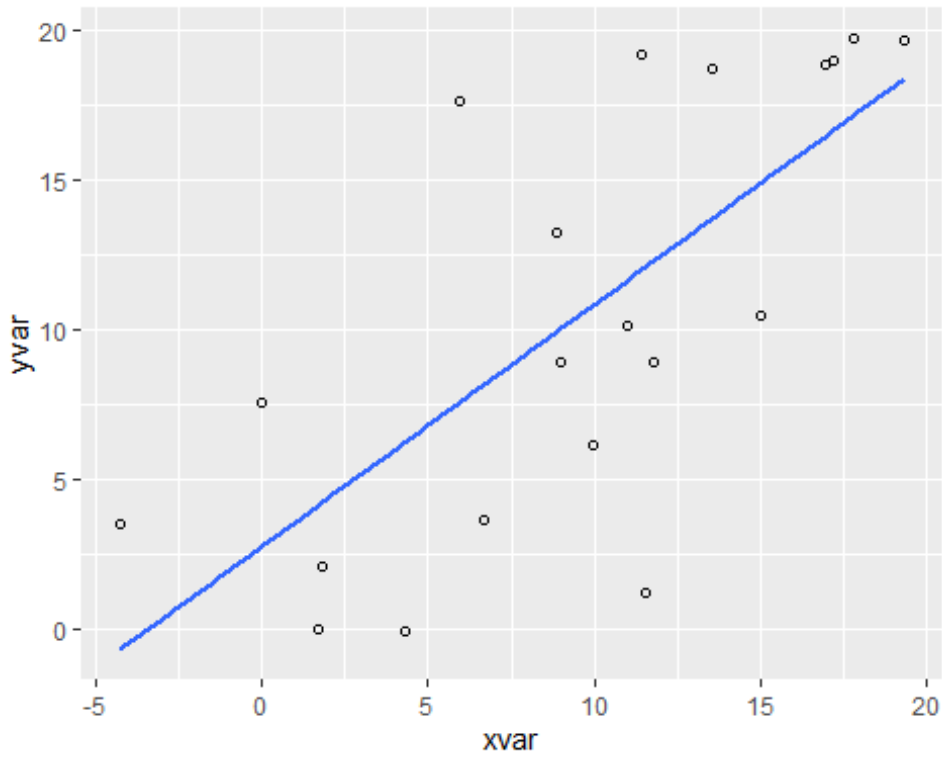
set.seed(955) is used to set a seed for the random number generator.

The **set.seed()** function sets the starting number used to generate a sequence of random numbers – it ensures that you get the same result if you start with that same seed each time you run the same process. If you omit this, or change the number, you will get a different data from the **rnorm()** function.

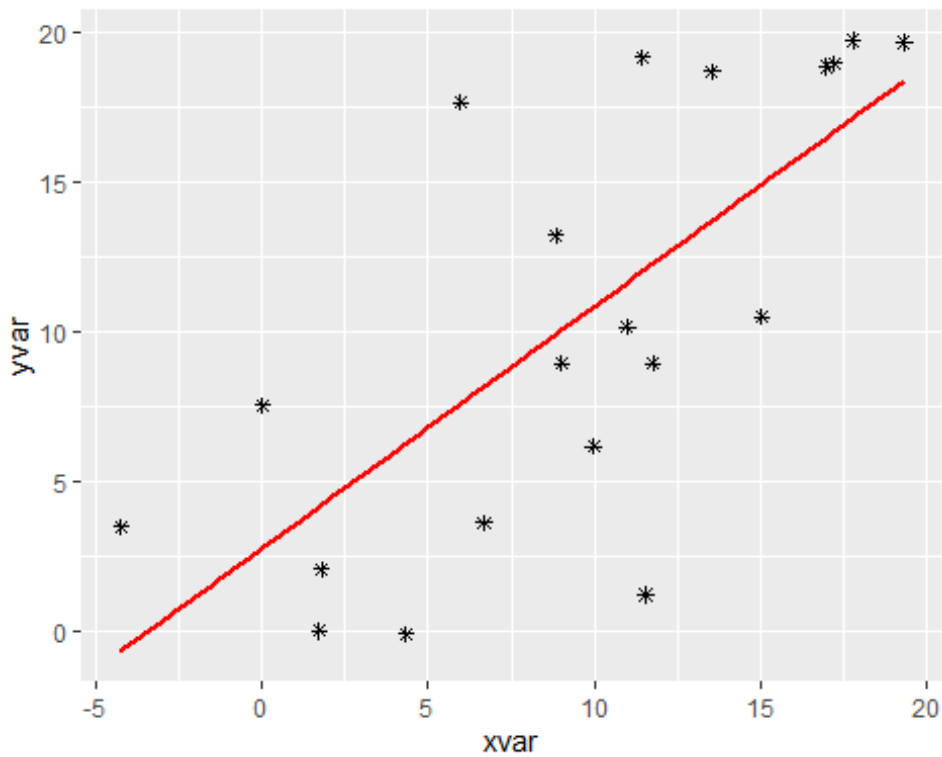
Question 1:

The R code below, is from [Cookbook for R](#). It creates a scatterplot with hollow circles and a linear regression line. But, the gray shaded confidence region is omitted.

```
ggplot(dat, aes(x=xvar, y=yvar)) +
  geom_point(shape=1) +      # Use hollow circles
  geom_smooth(method=lm,    # Add linear regression line
              se=FALSE)     # Don't add shaded confidence region
```



Your Turn: Change the code above to generate the graph below:

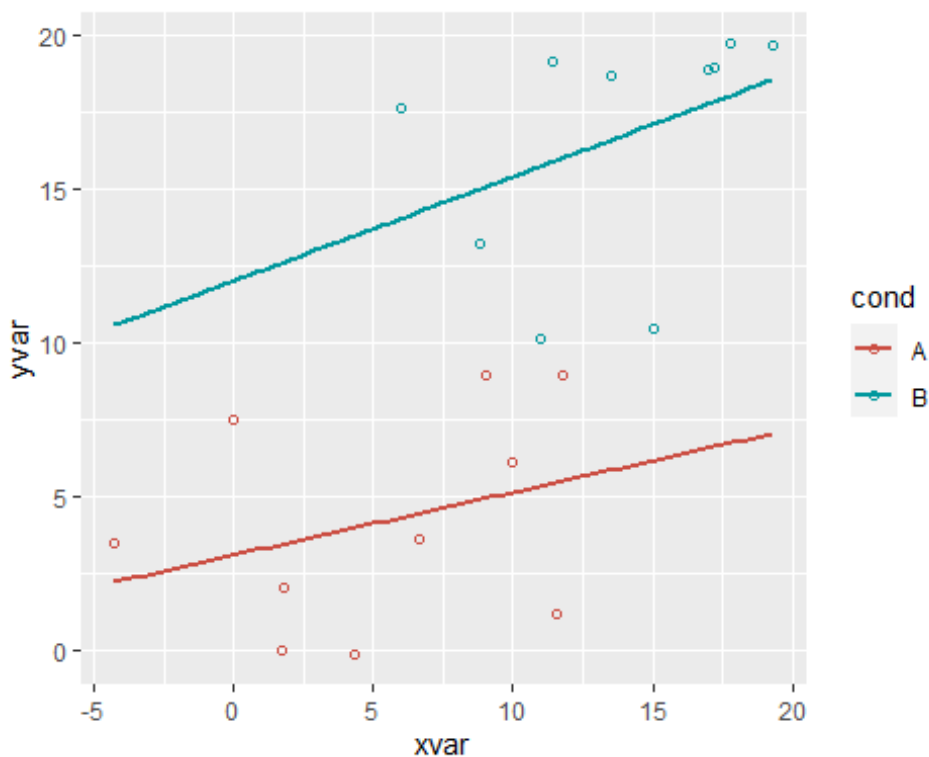


Please note that a nice resource for learning about point shapes is the [STHDA](#) post on [ggplot2 point shapes](#). Alboukadel Kassambara's site has many wonderful R and ggplot2 resources. I encourage you to check them out.

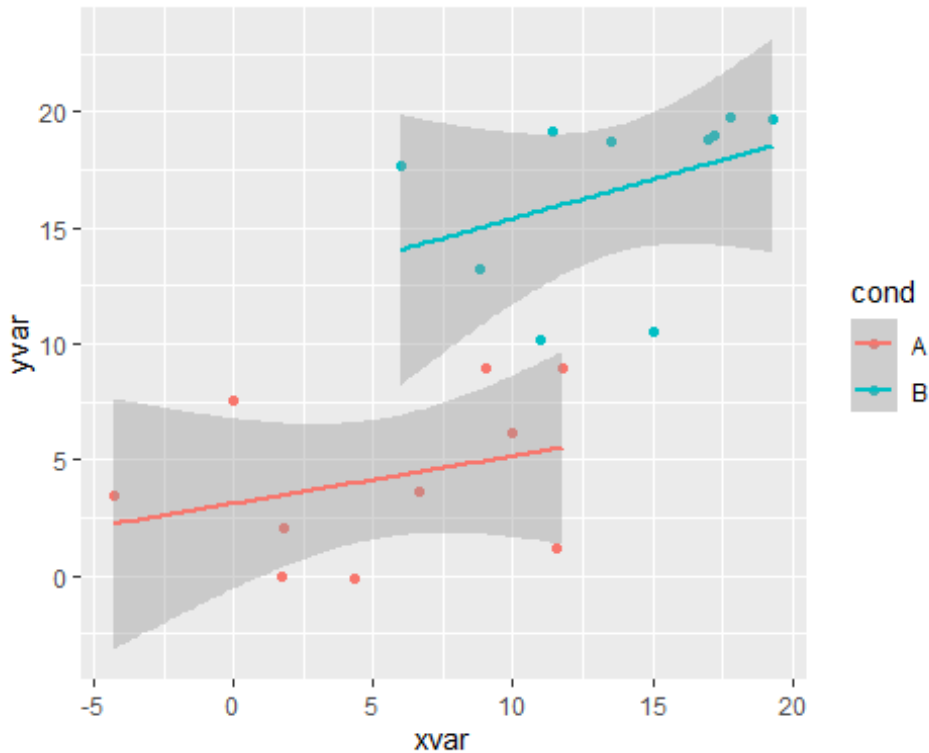
Question 2:

The R code below, is also from [Cookbook for R](#). It creates a scatterplot with hollow circles and two liner regression lines – one for each **cond** (i.e., A and B). Again, the gray shaded confidence region is omitted. The graph also uses slightly darker palette than normal.

```
# Extend the regression lines beyond the domain of the data  
ggplot(dat, aes(x=xvar, y=yvar, color=cond)) + geom_point(shape=1) +  
  scale_colour_hue(l=50) + # Use a slightly darker palette than normal  
  geom_smooth(method=lm, # Add linear regression lines  
              se=FALSE, # Don't add shaded confidence region  
              fullrange=TRUE) # Extend regression lines
```



Your Turn: Change the code above to generate the graph below:



Please note that the dots are *solid* and the shaded confidence regions *are present*.

Question 3

This next question uses an example from the [STHDA](#) post [ggplot2 barplots : Quick start guide - R software and data visualization](#)

Data

The dataframe below (**df**) is derived from the *ToothGrowth* data. The *ToothGrowth* data describes the effect of Vitamin C on tooth growth in Guinea pigs.

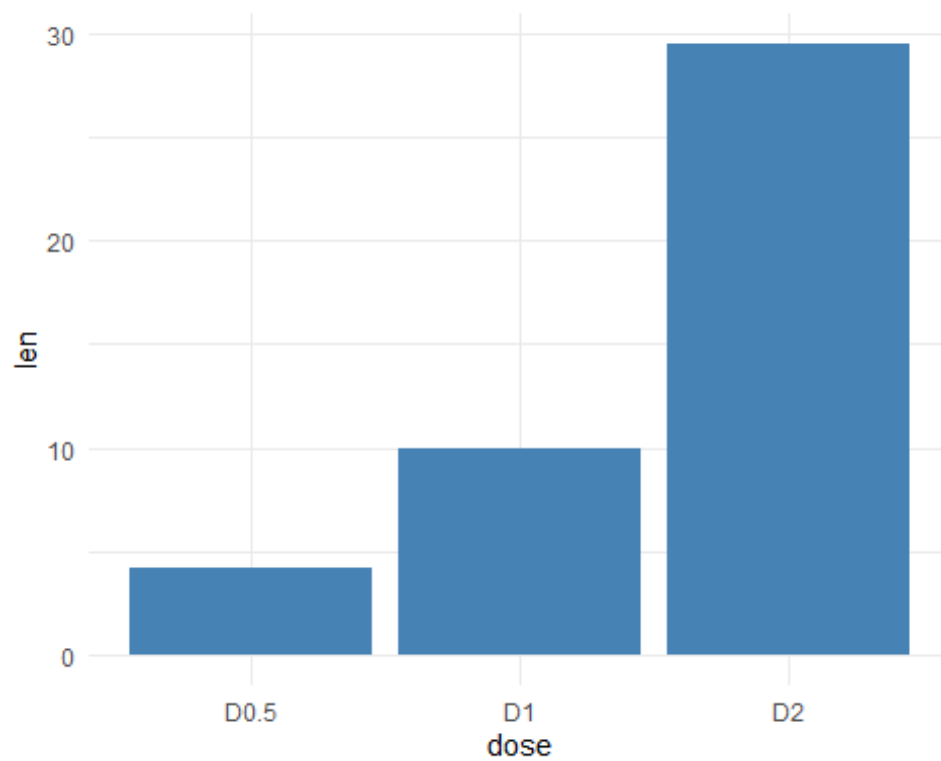
```
df <- data.frame(dose=c("D0.5", "D1", "D2"),
                 len=c(4.2, 10, 29.5))
head(df)
```

```
##   dose  len
## 1 D0.5  4.2
## 2  D1 10.0
## 3  D2 29.5
```

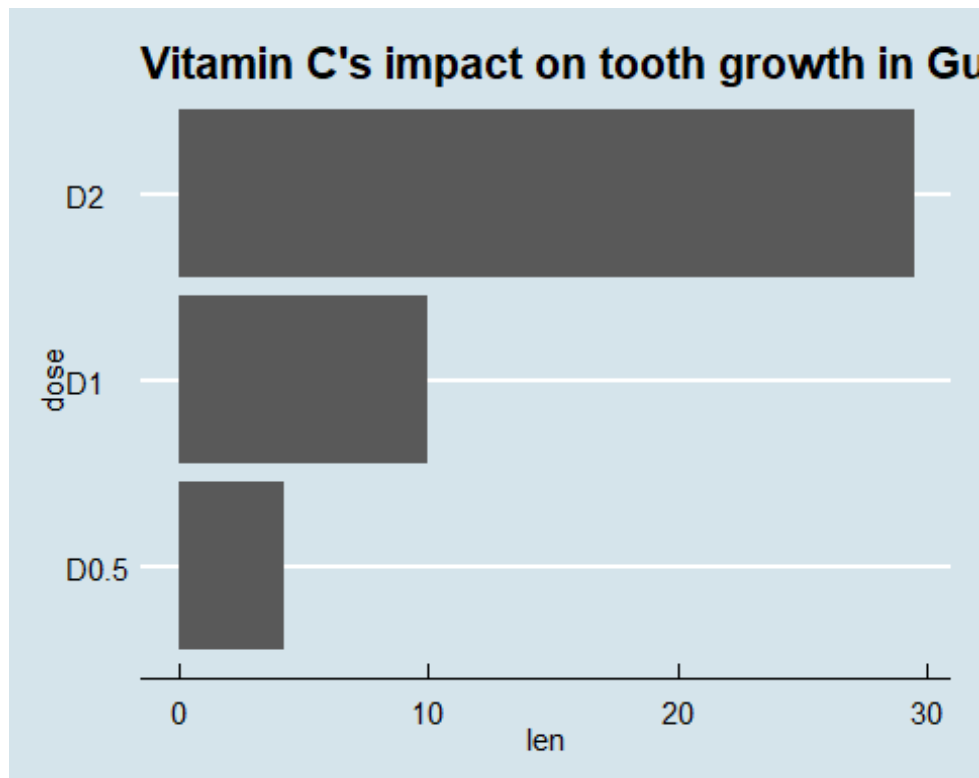
The graph below is a **Bar plot** with **steelblue** bars. It uses the **minimal theme**.

```
# Minimal theme + blue fill color
p<-ggplot(data=df, aes(x=dose, y=len)) +
  geom_bar(stat="identity", fill="steelblue")+
  theme_minimal()
```

```
theme_minimal()
p
```



Your Turn: Change the code above to generate the graph below:



Links that might be helpful for this question

[ggplot2 colors](#)
[ggplot2 themes](#)

The theme used is from the Economist

Question 4:

Data for question 4.

```
set.seed(1234)
dat <- data.frame(cond = factor(rep(c("A", "B"), each=200)),
                  rating = c(rnorm(200), rnorm(200, mean=.8)))
# View first few rows
head(dat)
```

	cond	rating
## 1	A	-1.2070657
## 2	A	0.2774292
## 3	A	1.0844412
## 4	A	-2.3456977
## 5	A	0.4291247
## 6	A	0.5060559

Above, **rnorm** generates a vector of 20 normally distributed random numbers.

rnorm can take up to three arguments:

n: the number of random variables to generate (**200**).

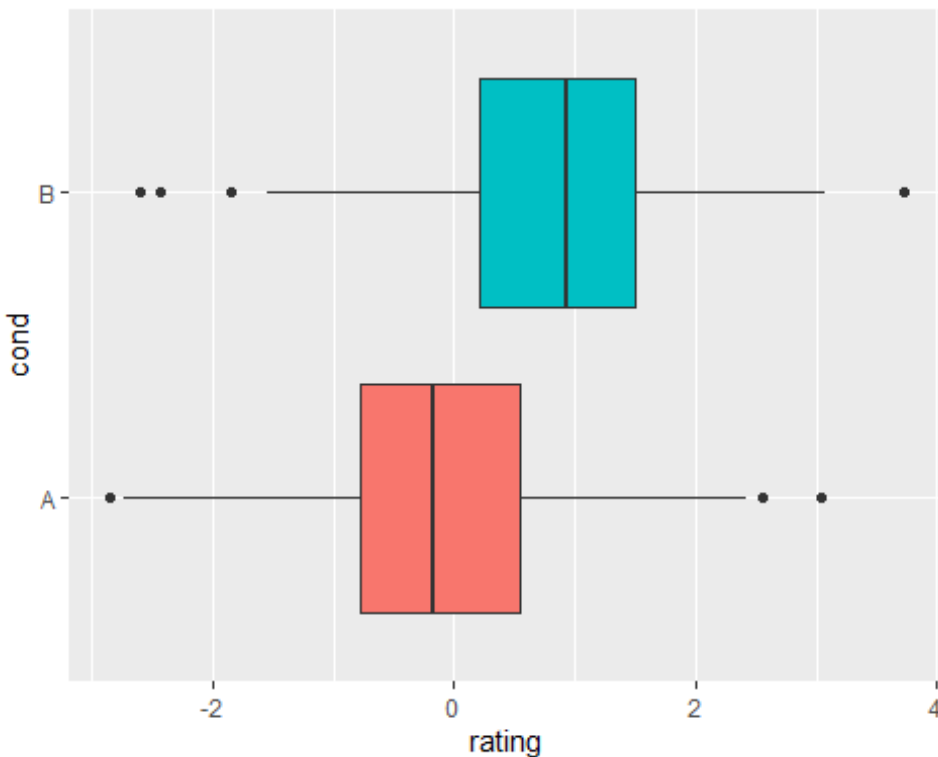
mean, if not specified defaults to 0 (**0.8**).

sd: the standard deviation (**not specifie**). If not specified, defaults to 1.

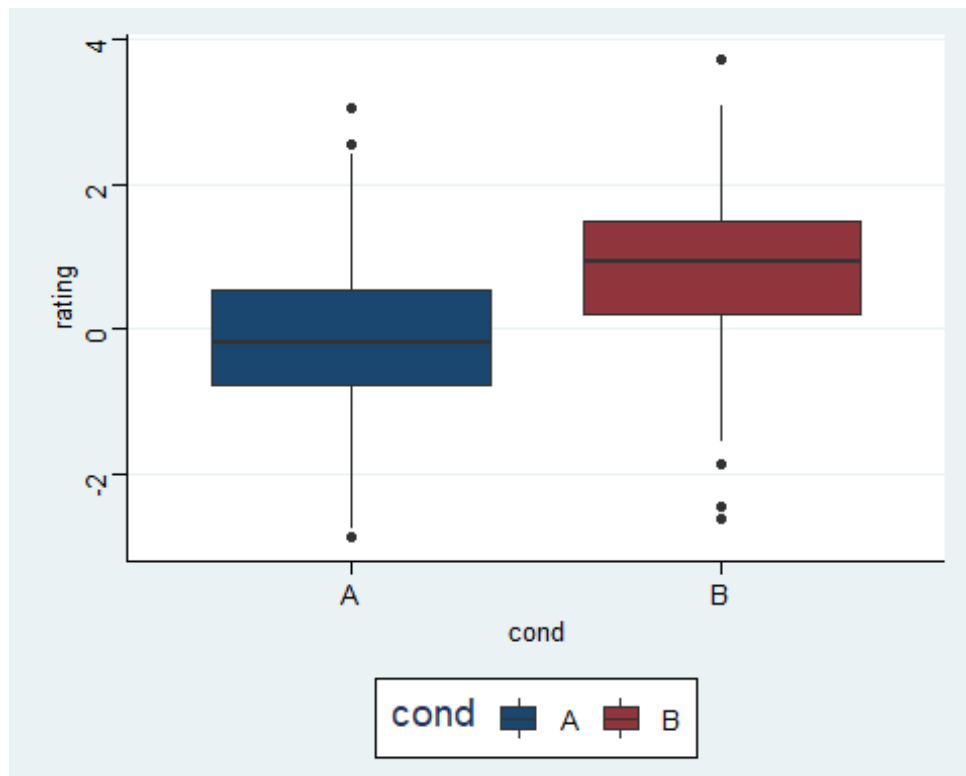
set.seed(1234) is used to set a seed for the random number generator.

The R code below, is from Cookbook for R's chapter on [Plotting distributions \(ggplot2\)](#). It creates a Boxplot with the conditions colored, the legend removed, and flipped axes.

```
ggplot(dat, aes(x=cond, y=rating, fill=cond)) + geom_boxplot() +  
  guides(fill=FALSE) + coord_flip()
```



Your Turn: Change the code above to generate the graph below:



The theme used is from Stata