

# WorkSheet-6

Plotting using ggplot() and plot()

## Worksheet for R Programming

### Instructions:

- Use RStudio or the RStudio Cloud accomplish this worksheet.
- Save the R script as *RWorksheet\_lastname#6.R*.
- On your own *GitHub repository*, push the R script, the Rmd file, as well as this pdf worksheet to the repo you have created before.
- Do not forget to comment your Git repo on our VLE
- Accomplish this worksheet by answering the questions being asked and writing the code manually.

Use the dataset **mpg**

```
library(ggplot2)

#to get the mpg dataset, load the ggplot package first
data(mpg)
as.data.frame(data(mpg)) #converting from list to data frame
```

```
## data(mpg)
## 1      mpg
```

A data frame with 234 rows and 11 variables:

```
#' \describe{
#'   \item{manufacturer}{manufacturer name}
#'   \item{model}{model name}
#'   \item{displ}{engine displacement, in litres}
#'   \item{year}{year of manufacture}
#'   \item{cyl}{number of cylinders}
```

```
#' \item{trans}{type of transmission}
#' \item{drv}{the type of drive train, where f = front-wheel drive, r = rear wheel drive}
#' \item{cty}{city miles per gallon}
#' \item{hwy}{highway miles per gallon}
#' \item{fl}{fuel type}
#' \item{class}{"type" of car}
#' }
"mpg"
```

```
str(mpg)
```

```
## tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
## $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
## $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
## $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
## $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
## $ drv         : chr [1:234] "f" "f" "f" "f" ...
## $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
## $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
## $ fl          : chr [1:234] "p" "p" "p" "p" ...
## $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

```
#use of glimpse() - much tidier compared to str()
library(dplyr) #glimpse() is a function under dplyr package
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

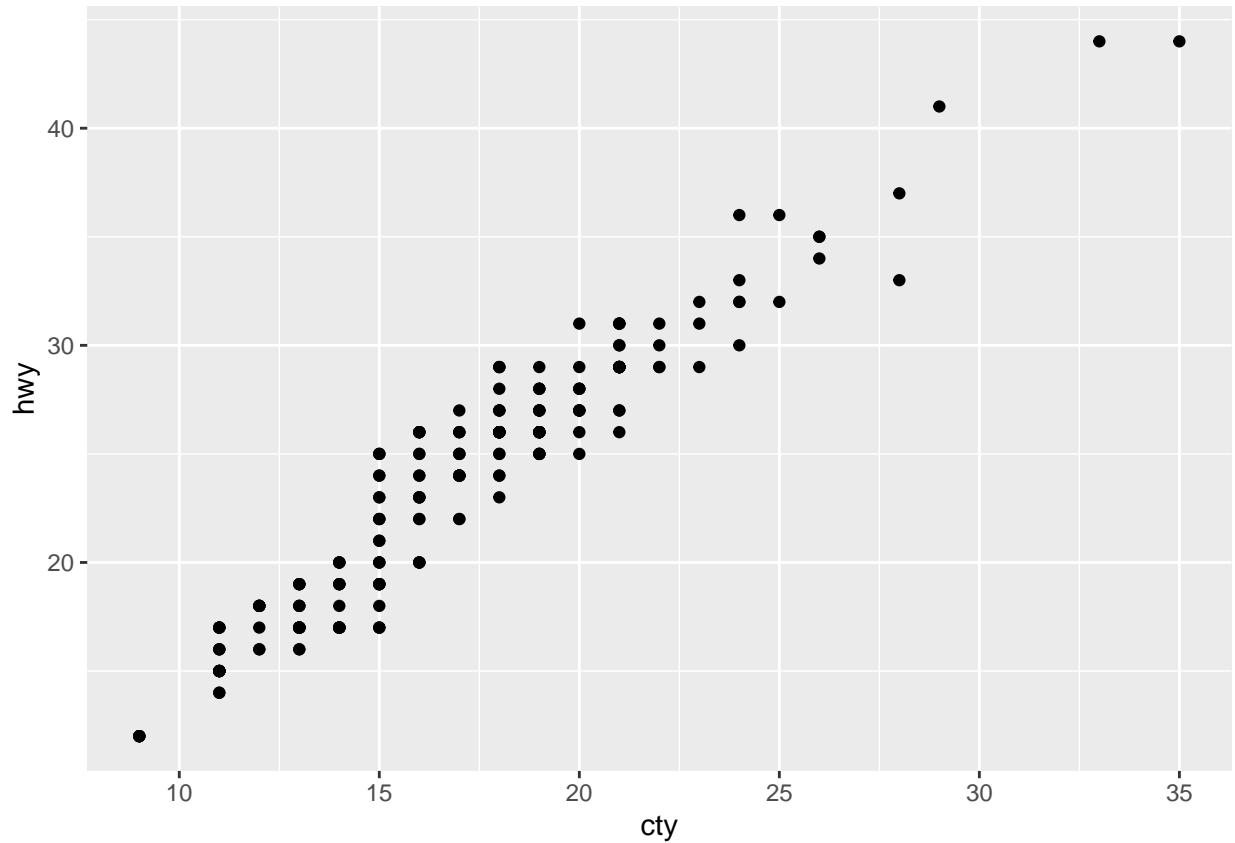
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
glimpse(mpg)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "~
## $ model        <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "~
## $ displ        <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~
## $ year         <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~
## $ cyl          <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 8, 8, ~
## $ trans        <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto~
## $ drv          <chr> "f", "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4~
## $ cty          <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~
## $ hwy          <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~
## $ fl           <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p~
## $ class        <chr> "compact", "compact", "compact", "compact", "compact", "c~
```

- Example. graph using ggplot()

```
ggplot(mpg, aes(cty, hwy)) +  
  geom_point()
```



1. How many columns are in `mpg` dataset? How about the number of rows? Show the codes and its result.
  
2. Which manufacturer has the most models in this data set? Which model has the most variations? Ans:
  - a. Group the manufacturers and find the unique models. Copy the codes and result.
  
  - b. Graph the result by using `plot()` and `ggplot()`. Write the codes and its result.

3. Same dataset will be used. You are going to show the relationship of the `model` and the `manufacturer`.

a. What does `ggplot(mpg, aes(model, manufacturer)) + geom_point()` show?

b. For you, is it useful? If not, how could you modify the data to make it more informative?

4. Using the pipe (`%>%`), group the `model` and get the number of cars per model. Show codes and its result.

a. Plot using the `geom_bar() + coord_flip()` just like what is shown below. Show codes and its result.

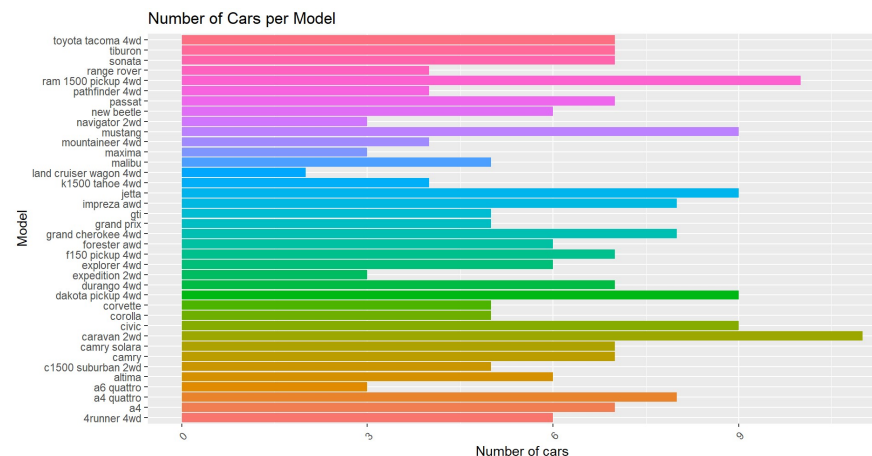


Figure 1: Car Models

b. Use only the top 20 observations. Show code and results.

5. Plot the relationship between `cyl` - *number of cylinders* and `displ` - *engine displacement* using `geom_point` with aesthetic `colour = engine displacement`. Title should be “**Relationship between No. of Cylinders and Engine Displacement**”.

a. Show the codes and its result.

b. How would you describe its relationship?

6. Get the total number of observations for `drv` - *type of drive train* (f = front-wheel drive, r = rear wheel drive, 4 = 4wd) and `class` - type of class (Example: suv, 2seater, etc.). Plot using the `geom_tile()` where the number of observations for `class` be used as a fill for aesthetics.

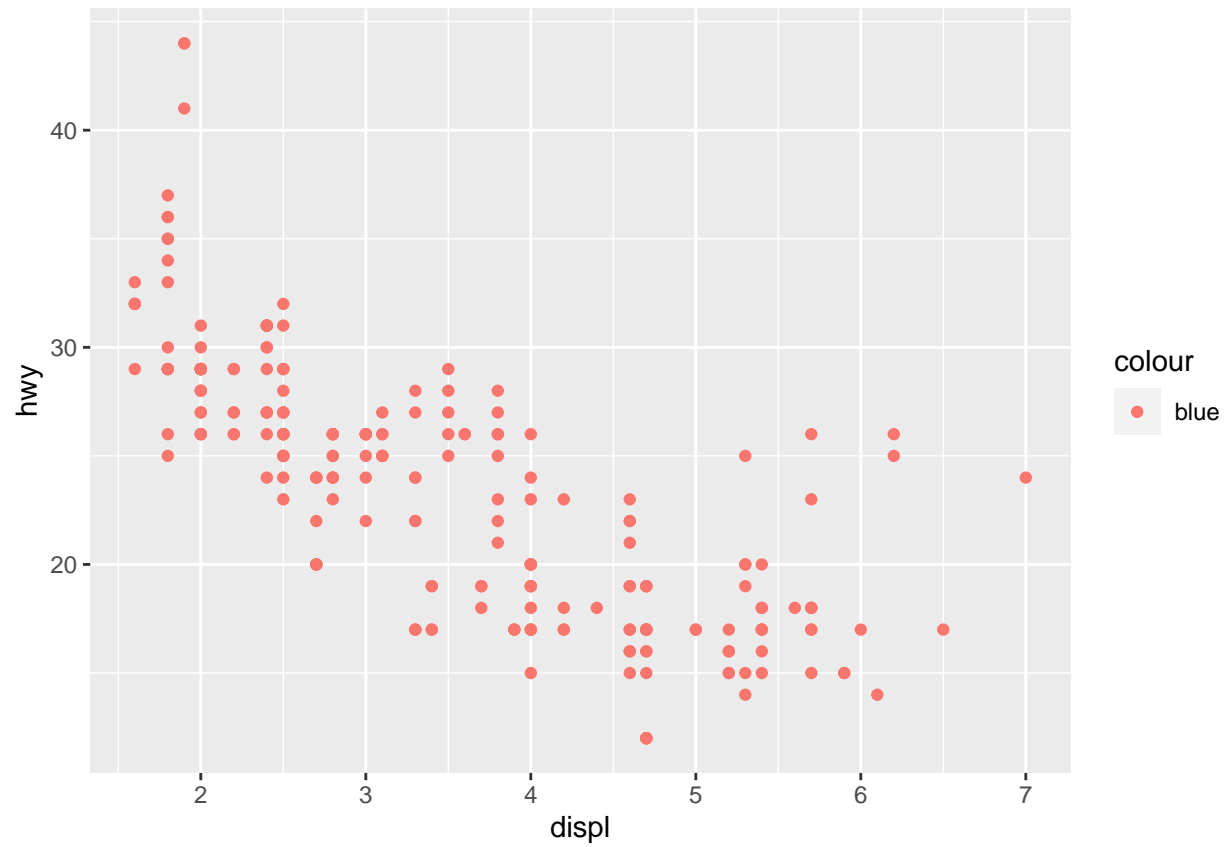
a. Show the codes and its result for the narrative in #6.

b. Interpret the result.

7. Discuss the difference between these codes. Its outputs for each are shown below.

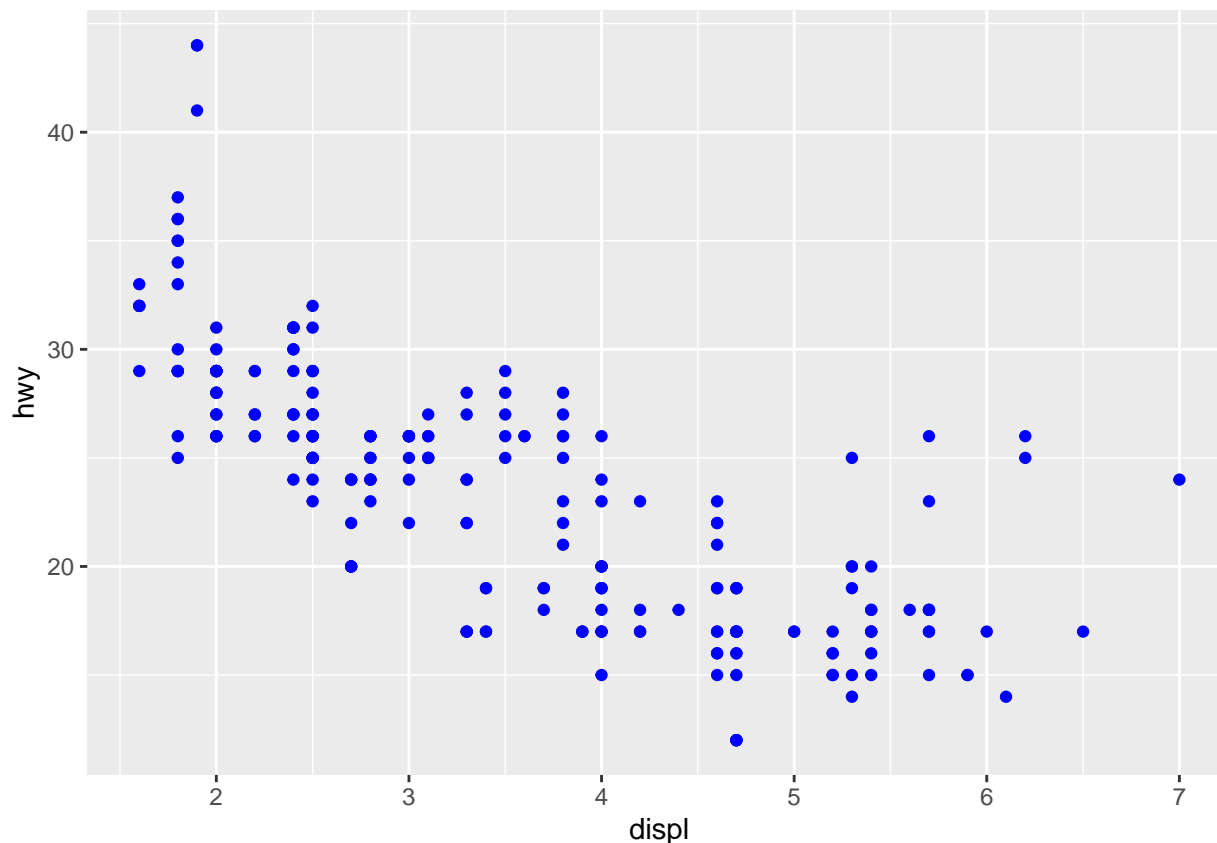
- Code #1

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, colour = "blue"))
```



+ Code #2

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy), colour = "blue")
```



8. Try to run the command `?mpg`. What is the result of this command?

a. Which variables from `mpg` dataset are categorical?

b. Which are continuous variables?

c. Plot the relationship between `displ` (engine displacement) and `hwy` (highway miles per gallon). Mapped it with a continuous variable you have identified in #5-b. What is its result? Why it produced such output?

9. Plot the relationship between `displ` (engine displacement) and `hwy` (highway miles per gallon) using `geom_point()`. Add a trend line over the existing plot using `geom_smooth()` with `se = FALSE`. Default method is *“loess”*.



10. Using the relationship of `displ` and `hwy`, add a trend line over existing plot. Set the `se = FALSE` to remove the confidence interval and `method = lm` to check for linear modeling.