Intro to R

Edwin Lo

Aug 26, 2019

R Markdown

Cars summary

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

The dataset/data frame cars is in the base library called "datasets" that comes with R. There are a lot more datasets to explore there.

Each code chunk can have R-markdown settings of echo=FALSE, eval=TRUE, results='hide', message=FALSE, warning=FALSE, etc. OR just include=FALSE will set all those options to FALSE (except eval=TRUE is the default).

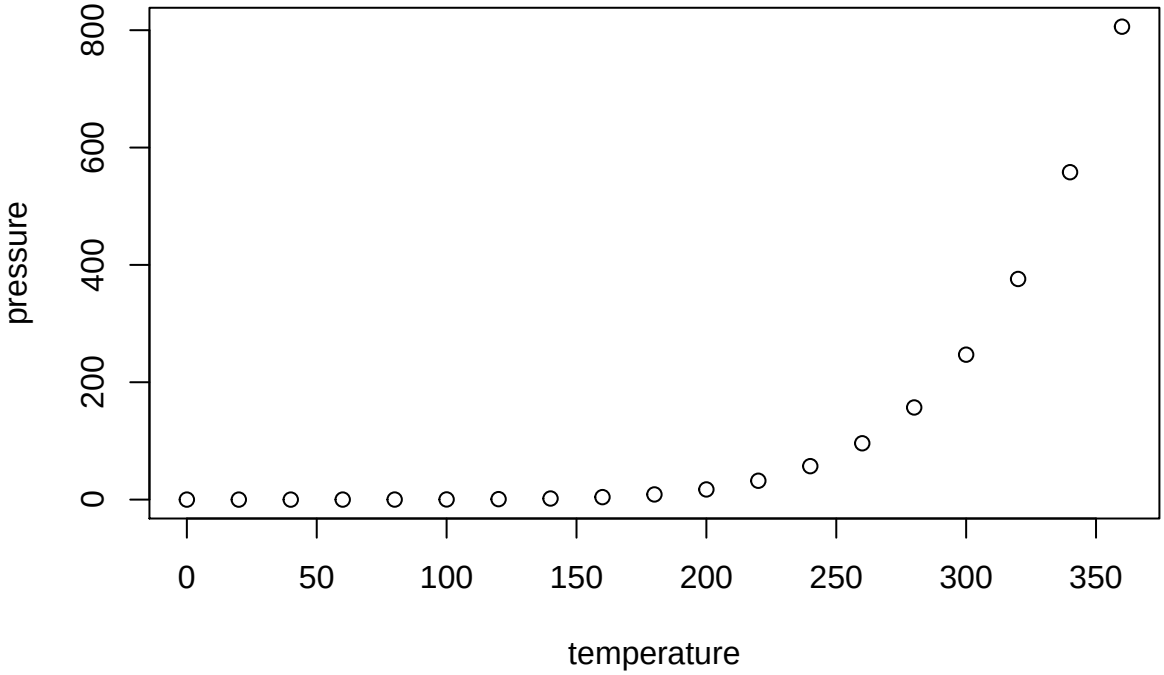
Including Plots

You can also embed plots, for example: First, try TRUE

```
summary(pressure)
```

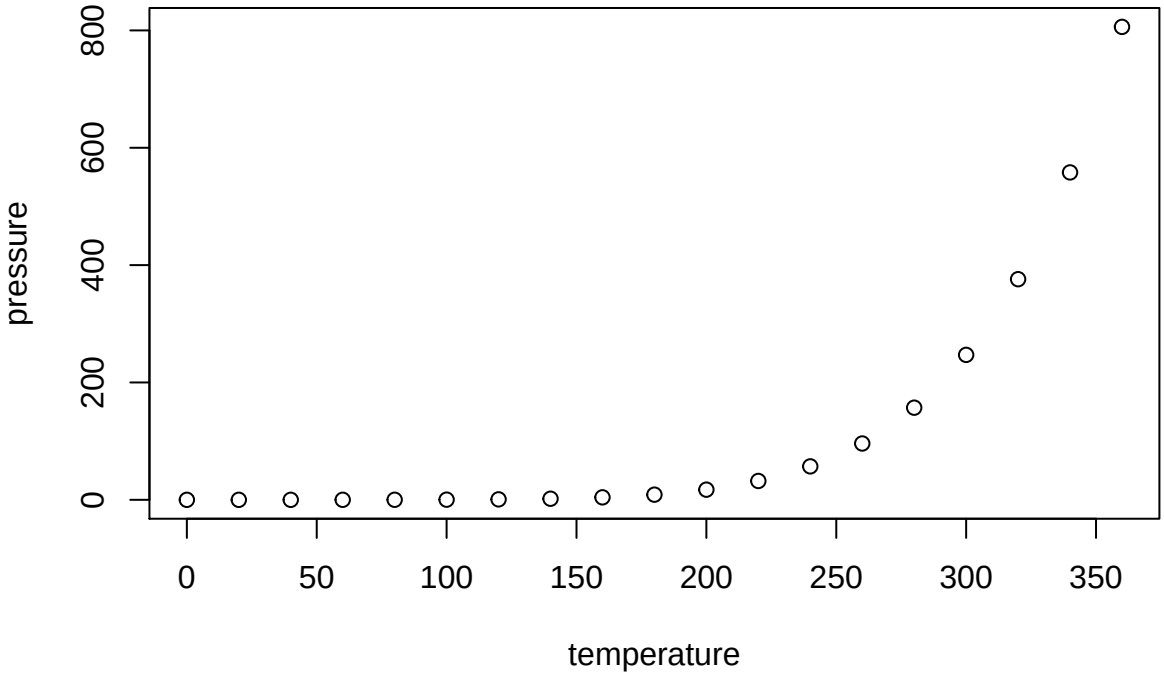
```
##   temperature     pressure      
##  Min.   :  0   Min.   :  0.0002  
##  1st Qu.: 90   1st Qu.:  0.1800  
##  Median :180   Median :  8.8000  
##  Mean   :180   Mean   :124.3367  
##  3rd Qu.:270   3rd Qu.:126.5000  
##  Max.   :360   Max.   :806.0000
```

```
plot(pressure)
```



Next, try echo=FALSE

```
##   temperature     pressure
##  Min.   :  0   Min.   :  0.0002
##  1st Qu.: 90   1st Qu.:  0.1800
##  Median :180   Median :  8.8000
##  Mean   :180   Mean   :124.3367
##  3rd Qu.:270   3rd Qu.:126.5000
##  Max.   :360   Max.   :806.0000
```



Note that the "echo = FALSE" parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Then we try include=FALSE

Plot is a general function in the base library to make plots. It can make different kinds of plots, depending on the dataframe you put in.

Header line (h4, bold faced)

Somewhat Bigger Header (h3)

Slightly Bigger Header (h2)

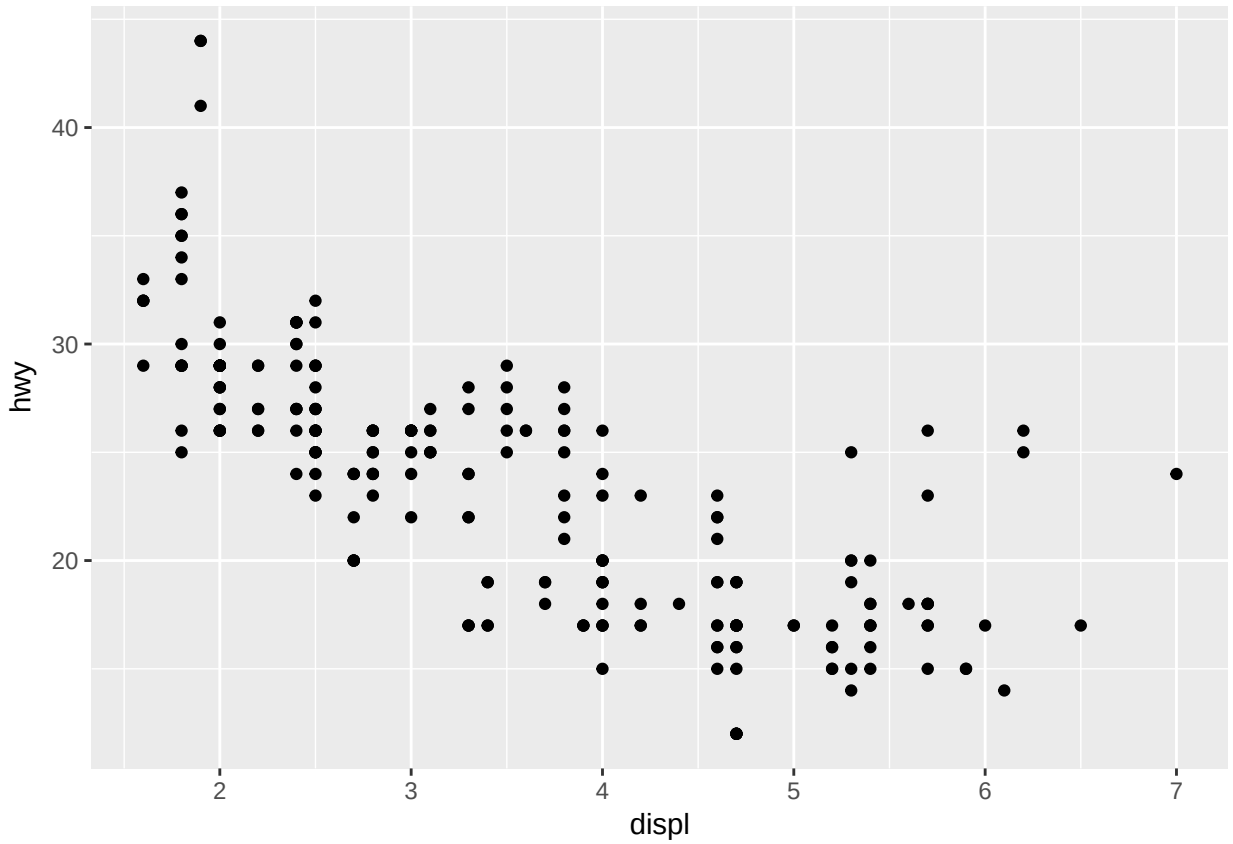
Biggest Header (h1): What is a package?

You can also include **in-line** R calculations like this: the average pressure is 124.3367053.

```
loadPkg("ggplot2") # you can use library with or without double quotes. But install.packages requires d
# edit(mpg)
# ?mpg # same as help(mpg)
head(mpg)
```

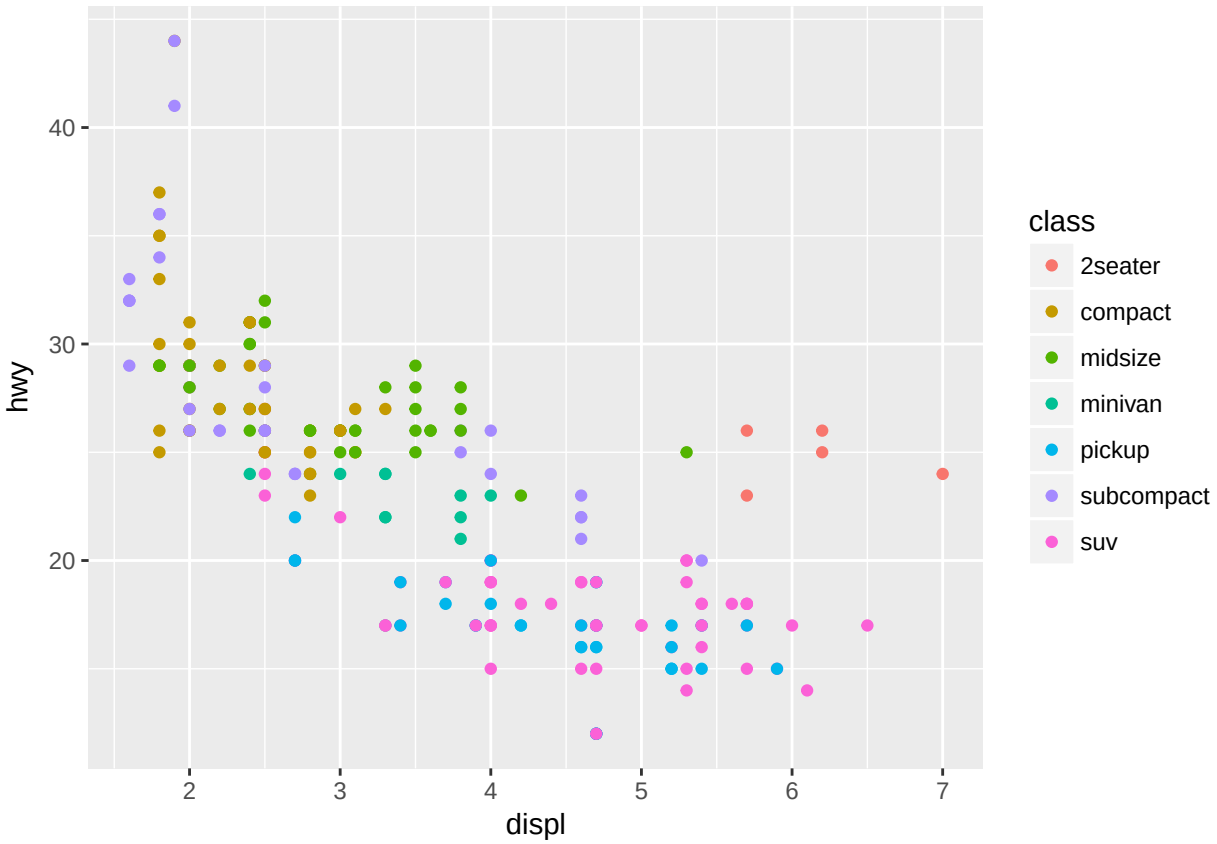
```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans  drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>  <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(~ f        18    29 p     comp~
## 2 audi         a4      1.8  1999     4 manua~ f        21    29 p     comp~
## 3 audi         a4      2    2008     4 manua~ f        20    31 p     comp~
## 4 audi         a4      2    2008     4 auto(~ f        21    30 p     comp~
## 5 audi         a4      2.8  1999     6 auto(~ f        16    26 p     comp~
## 6 audi         a4      2.8  1999     6 manua~ f        18    26 p     comp~
```

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



Now we can also add a color into the mix with a simple adjustment to the code

```
ggplot(data=mpg)+
  geom_point(mapping = aes(x=displ, y=hwy, color=class))
```



So how do we get data into R: First we have to know where R is pulling our data from, we can do this by determing the location of our current working directy

```
getwd()
```

```
## [1] "/Users/Jeffrey/Documents/GitHub/GWU_classes/DATS_6101_IntroDS/Class01_Intro"
```

If you want to move your working directory you can use the setwd("Insert File Location Here")

```
Admit = data.frame(read.csv("LogRegAdmit.csv"))
head(Admit)
```

```
##   admit gre  gpa rank
## 1     0 380 3.61    3
## 2     1 660 3.67    3
## 3     1 800 4.00    1
## 4     1 640 3.19    4
## 5     0 520 2.93    4
## 6     1 760 3.00    2
```

```
# loadPkg("ggplot2") # already loaded before
head(Admit)
```

```
##   admit gre  gpa rank
## 1     0 380 3.61    3
## 2     1 660 3.67    3
## 3     1 800 4.00    1
## 4     1 640 3.19    4
## 5     0 520 2.93    4
## 6     1 760 3.00    2
```

```
loadPkg("tidyverse")
```

```
## Error: package or namespace load failed for 'tidyverse' in loadNamespace(j <- i[[1L]], c(lib.loc, .li
##  there is no package called 'broom'
```

```
## also installing the dependencies 'broom', 'dbplyr', 'haven', 'modelr', 'feather'
```

```
##
##   There are binary versions available but the source versions are
##   later:
##         binary source needs_compilation
## broom    0.4.5  0.5.2             FALSE
## dbplyr   1.3.0  1.4.2             FALSE
## haven    2.1.0  2.1.1              TRUE
## modelr   0.1.4  0.1.5             FALSE
## feather  0.3.3  0.3.4              TRUE
##
##
## The downloaded binary packages are in
##  /var/folders/fc/4s5ks50s5453qv5zy8s1f5mh0000gn/T//Rtmp0VrcE6/downloaded_packages
```

```
## installing the source packages 'broom', 'dbplyr', 'haven', 'modelr', 'feather'
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages ---------------------------------------------------- tidyverse 1.2.1 --
```

```
## v tibble  2.1.3     v purrr   0.3.2
## v tidyr   0.8.3     v dplyr   0.8.3
## v readr   1.1.1     v stringr 1.3.0
## v tibble  2.1.3     v forcats 0.4.0
```

```
## Warning: package 'tidyr' was built under R version 3.4.4
```

```
## Warning: package 'purrr' was built under R version 3.4.4
```

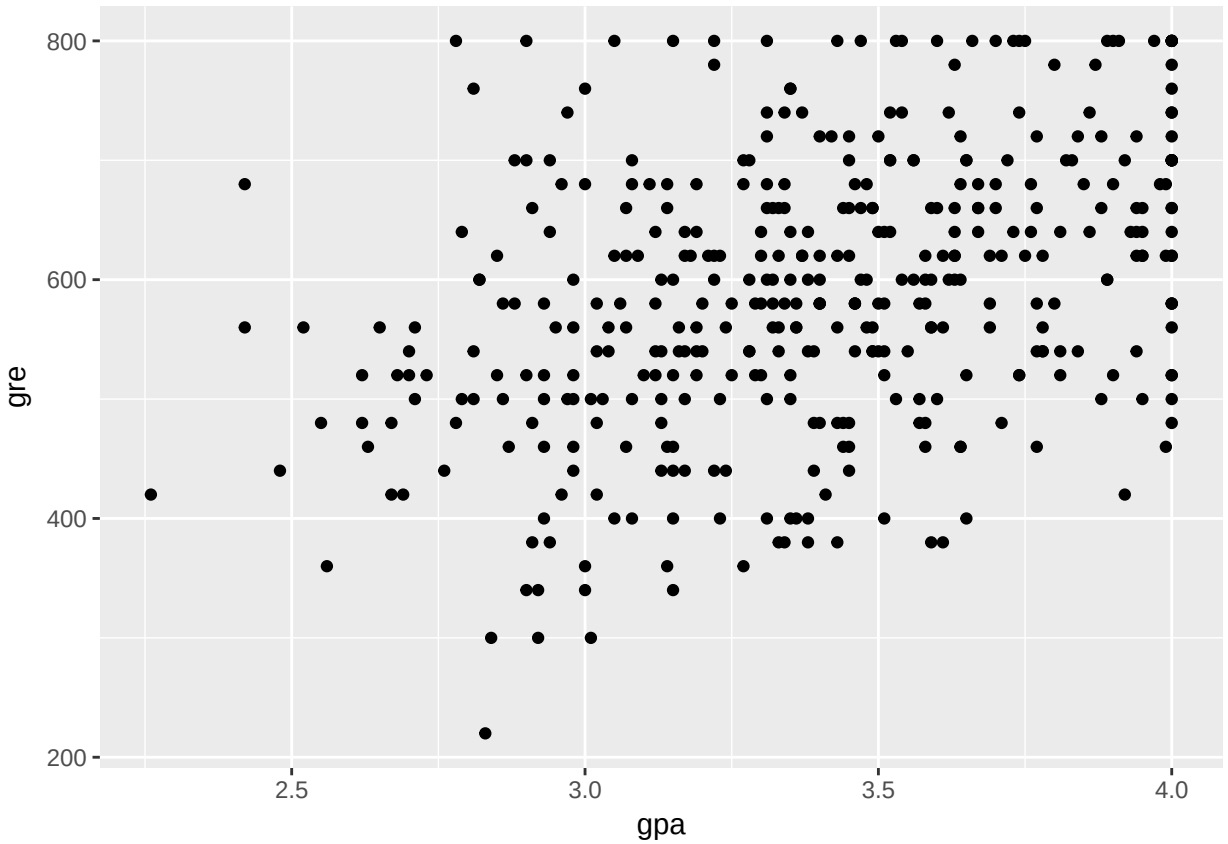
```
## Warning: package 'forcats' was built under R version 3.4.4
```

```
## -- Conflicts ------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
glimpse(Admit)
```

```
## Observations: 400
## Variables: 4
## $ admit <int> 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,...
## $ gre   <int> 380, 660, 800, 640, 520, 760, 560, 400, 540, 700, 800, 4...
## $ gpa   <dbl> 3.61, 3.67, 4.00, 3.19, 2.93, 3.00, 2.98, 3.08, 3.39, 3....
## $ rank  <int> 3, 3, 1, 4, 4, 2, 1, 2, 3, 2, 4, 1, 1, 2, 1, 3, 4, 3, 2,...
```

```
ggplot(data=Admit)+
  geom_point(mapping = aes(x=gpa, y=gre))
```




```
# detach("package:ggplot2", unload = T) # consider detach packages no longer use to free up system memo
```

As you might now notice, install.packages does not work well with knitr. You will need to comment out these lines before knitting. Other functions such as "View", "edit", which shifts the focus to another window/tab also do not work well with knitting. You will also need to comment out those lines as well. Note that without double blank space in the previous line, is different from here, with 2 spaces at the end.
Can you see the difference?

With or without 2 spaces at the end, if there is a blank line, then it will always be a new paragraph.