Curso: R para Data Science

**1. Introdução**

R é um ambiente integrado para manipulação de dados, cálculo e exibição gráfica. Ele está disponível gratuitamente e compila/roda em uma grande variedade de sistemas operacionais, como UNIX , Windows e MacOS.

**2. Download R**

Para fazer o download do R em seu computador basta seguir as instruções do link:

<https://cran.r-project.org/bin/windows/base/>

**3. RStudio**

Rstudio é um conjunto de ferramentas integradas, destinadas a ajudá-lo a ser mais produtivo com R a partir da sua interface. Ele inclui um console, um editor para realce da sintaxe que suporta a execução direta do código, bem como ferramentas para plotar, acessar histórico, depuração e gerenciamento de espaço de trabalho. Você pode fazer o download em seu computador seguindo as instruções do link:

<https://www.rstudio.com/products/rstudio/download/>

#Pacote que precisara instalar. Um pacote R eh uma colecao de funcoes.

#usaremos o pacote tidyverse. Instalando: na linha de comando digitar:

install.packages("tidyverse")

#Uma vez instalado você poderá carregá-lo com a funcão:

library(tidyverse)

#outros pacotes:Estes pacotes fornecem dados sobre linhas aereas,

#desenvolvimento mundial e baseball

install.packages(c("nycflights13","gapminder","Lahman"))

#ou um por um:

install.packages("nycflights13")

install.packages("gapminder")

install.packages("Lahman")

#Uma vez instalado você poderá carregá-lo:

library(nycflights13)

library(gapminder)

library(Lahman)

#mostra em que dir esta

getwd()

#muda de diretorio

setwd("C:/Users/mfariajr/Desktop/dataScience")

#conteudo diretorio

dir()

# exemplo de dados encontrado no ggplot: (ggplot::mpg)

# o mpg é uma base americana de 38 modelos de carros

# agencia de protecao ambiental dos Est. Und.

mpg

## A tibble: 234 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(l5) f 18 29 p compact

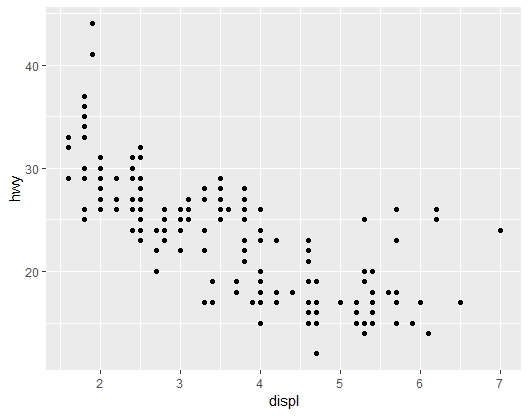
#2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compact

#3 audi a4 2 2008 4 manual(m6) f 20 31 p compact

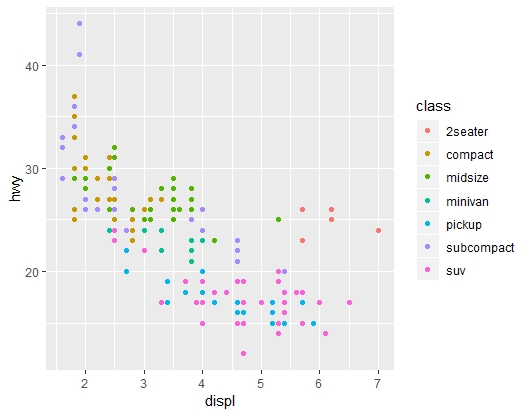
#4 audi a4 2 2008 4 auto(av) f 21 30 p compact

#grafico ggplot tamanho do motor: eixo-x (displ) e eficiência #eixo-y (hwy)

ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy))



> ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy, color= class))



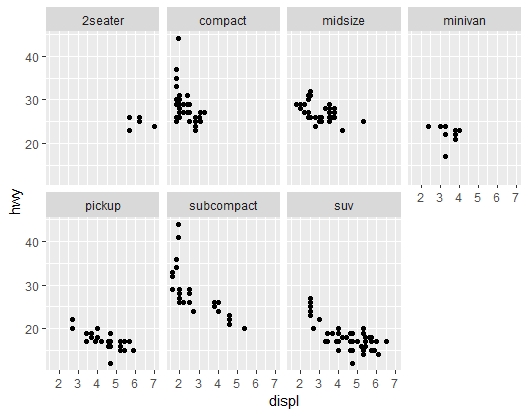
ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy, shape= class))

ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy, size= class))

ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy), color= "blue")

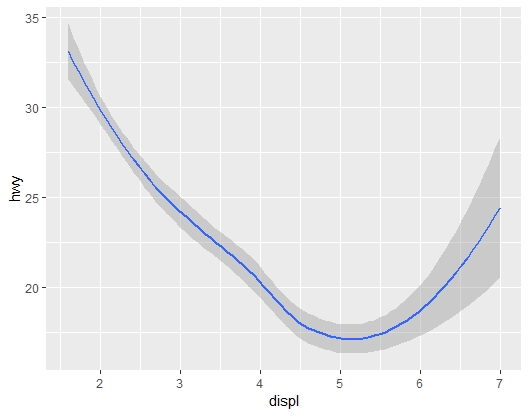
ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy , alpha = class), color= "blue")

ggplot(data = mpg)+geom\_point(mapping= aes(x=displ, y=hwy)) + facet\_wrap(~ class, nrow = 2)



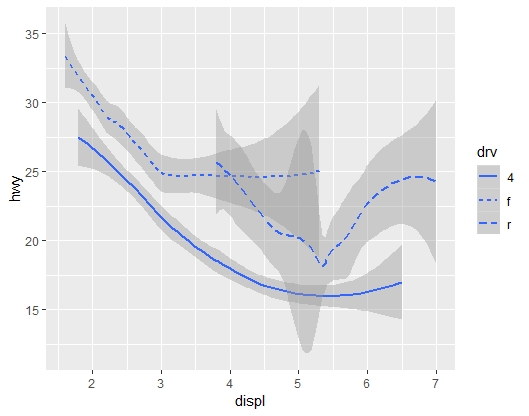
# uso de geom\_smooth

ggplot(data = mpg)+geom\_smooth(mapping= aes(x=displ, y=hwy))



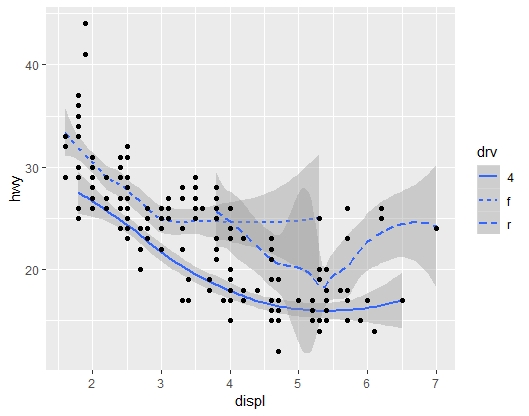
#uso de geom\_smooth linetype

ggplot(data = mpg)+geom\_smooth(mapping= aes(x=displ, y=hwy, linetype = drv))



#uso de geom\_smooth + geom\_point

ggplot(data = mpg)+geom\_smooth(mapping= aes(x=displ, y=hwy, linetype = drv))+geom\_point(mapping= aes(x=displ, y=hwy))



# média mpg 9 é endereço hwy na lista e os dados 1 até 234

mean(mpg[[9]][1:234])

[1] 23.44017

# desvio padrão hwy na lista valore 1 até 234

sd(mpg[[9]][1:234])

[1] 5.954643

#z-score

zsc<-(mpg[[9]][1:234]-mean(mpg[[9]][1:234]))/sd(mpg[[9]][1:234])

#erro err:

err= 1.96\* sd(mpg[[9]][1:234])/ sqrt(234)

#intervalo de confianca para 95%

mean(mpg[[9]][1:234])+err

#[1] 24.20313

mean(mpg[[9]][1:234])-err

#[1] 22.67721

|  |
| --- |
| order(zsc,decreasing = T)  # [1] 213 222 223 197 106 107  sort(zsc,decreasing = T)  # [1] 3.45273890 3.45273890 2.94893040 2.27718573 2.10924956 2.10924956 |
|  |
| |  | | --- | |  | |

#cricao de um vetor com expressao do gene A em 20 amostras

vec <- c(84.334,90.455,34.989,43.222,23,25.424,0.568,0.349,45.984,62.847,17.233,12.085,72.692,4.297,36.592,49.849,96.765,19.364,28.428,1.376)

#observar o vetor vec

vec

#verificar modo/tipo do vetor vec

mode(vec)

#verificar tamanho do vetor vec

length(vec)

#verificar estrutura do vetor vec

str(vec)

#verificar a classe do vetor vec

class(vec)

#observar 6 primeiros elementos do objeto

head(vec)

#ordenar o vetor vec criado

sort( vec , decreasing = TRUE )

#ordenar o vetor vec criado

order(vec)

order(vec, decreasing = T)

#arrendondar valores do vec

round(vec)

#selecionar quinto elemento do vetor vec

vec[5]

#selecionar do segundo ao decimo elemento do vetor vec

vec[2:10]

#selecionar o segundo e o quinto elemento do vetor vec criado

vec[c(2,5)]

#acessar quais variaveis estao salvas no ambiente R

ls()