Caderno de R

Lucas Ketzer dos Reis

### AULA 2

**OBS: nao foram utilizados acentos ortograficos para evitar erros ou caracteres nao reconhecidos no envio das tarefas**

## GRAFICOS COM GGPLOT2

No vídeo, ha uma breve explicacao da base de dados que sera utilizada e como ela sera manipulada usando dplyr para a construcao dos graficos.

Em seguida, ha uma explicacao da logica de camadas do ggplot2. Essa logica consiste no fato de que graficos em R sao construidos em camadas, com uma camada “em branco” sendo inicializada por um objeto ggplot que vai sendo incrementado com as camadas de dados (que definem quais dados serao utilizados na construcao do grafico), estetica (que definem quais variaveis vao em quais eixos, se estao agrupadas por alguma outra variavel ou nao, etc) e geometrica/geom, que efetivamente define qual sera o tipo de grafico.

Vale notar que diferentes geoms podem ser combinados em um mesmo grafico, e que a estetica pode ser declarada dentro de cada geom ou dentro do proprio objeto ggplot que inicializa o grafico.

O primeiro grafico apresentado eh o grafico de pontos, que comumente eh utilizado para fazer graficos de dispersao:

library(tidyverse)  
  
base <- read.csv("LucroFilmes.csv", sep = ";")  
  
base <- base %>%  
 mutate(DataLancamento = as.Date(DataLancamento, format = "%m/%d/%Y")) %>%  
 mutate(  
 AnoLancamento = format(DataLancamento, "%Y"),  
 MesLancamento = format(DataLancamento, "%m"),  
 LucroLocal = as.numeric(LucroLocal),   
 LucroMundial = as.numeric(LucroMundial)  
 )  
  
ggplot(base) +  
 geom\_point(aes(LucroLocal, LucroMundial))

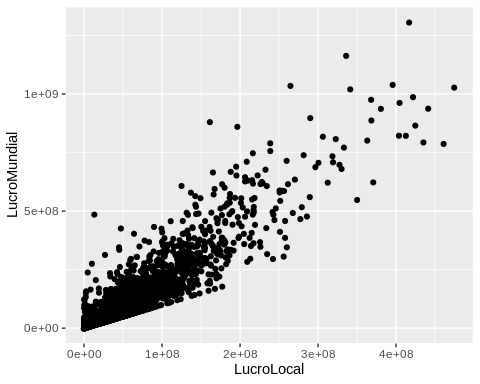
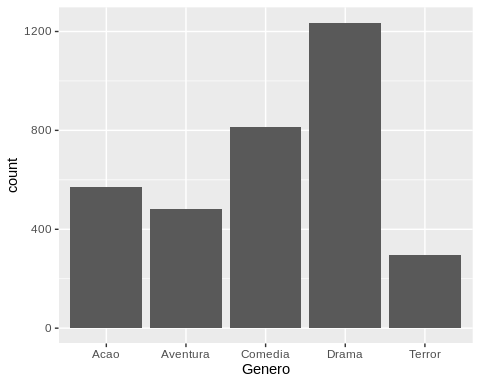


Grafico de pontos/dispersao

Em seguida, sao apresentados os graficos:

Histograma: Utilizado para avaliar distribuicoes de frequencias, a base do retangulo representa a faixa/classe da variavel estudada e a altura a quantidade do conjunto que se encontra nessa classe/faixa.

ggplot(base) +  
 geom\_histogram(aes(Genero), stat = "count")



Histograma

Grafico de barras: Um tipo de grafico bastante comum, onde as barras representam proporcionalmente os valores apresentados. Pode ser utilizado tanto com valores absolutos quanto com percentuais, podendo tambem ser agrupado/empilhado.

ggplot(base) +  
 geom\_bar(aes(Distribuidor), stat = "count")

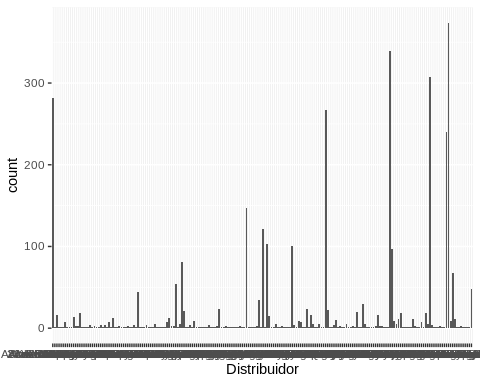
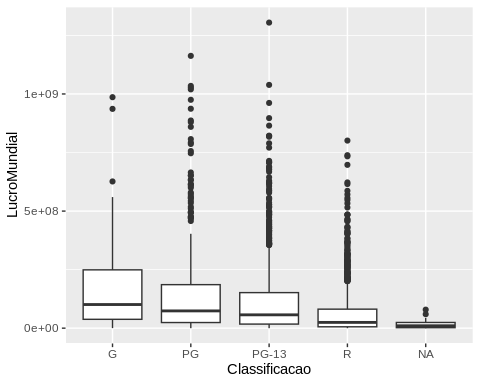


Grafico de barras

Boxplot/Diagrama de caixa: Um tipo de grafico bastante util em que se pode visualizar a distribuicao de um dado - o grafico indica os valores minimos e maximos, mediana e os percentis 25 e 75, alem de outliers.

ggplot(base) +   
 geom\_boxplot(aes(Classificacao, LucroMundial))



Boxplot/Diagrama de caixa

Grafico de linha: Normalmente utilizado para demonstrar a evolucao de alguma variavel ao longo do tempo.

ggplot(base) +   
 geom\_line(aes(AnoLancamento, group = 1), stat = "count")

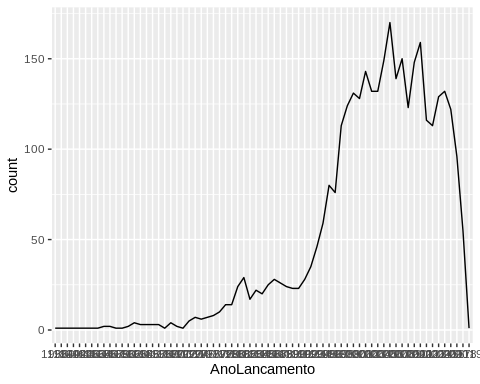


Grafico de linha

O restante do video eh dedicado a personalizacao de parametros do grafico, como tamanho, cor, preenchimento, titulo, legendas etc. Ha tambem uma secao demonstrando como combinar geoms, com esteticas especificas em cada um deles - tambem eh demonstrado como filtrar dados especificos antes da geracao do grafico e como adicionar estatisticas sumarias.

A fim de manter a brevidade, todos esses parametros serao demonstrados em um unico grafico de autoria propia:

base %>%  
 filter(as.numeric(AnoLancamento) >= 2000) %>%  
 group\_by(Distribuidor) %>%  
 mutate(total\_filmes = n()) %>% # usando esses dados p/ ver as dez maiores, as demais, serao agrupadas em outros  
 ungroup() %>%  
 mutate(distribuidor\_agrupado = ifelse(  
 total\_filmes < 68, "Outros", Distribuidor  
 )) %>%  
 group\_by(distribuidor\_agrupado, Genero) %>%  
 summarise(lucro\_medio\_por\_genero = mean(LucroMundial, na.rm = TRUE)) %>%  
 mutate(lucro\_medio\_por\_genero = round(lucro\_medio\_por\_genero / 100000)) %>%  
 ungroup() %>%  
 ggplot() +  
 geom\_point(aes(distribuidor\_agrupado, lucro\_medio\_por\_genero, color = Genero)) +  
 geom\_text(aes(distribuidor\_agrupado, lucro\_medio\_por\_genero, label = lucro\_medio\_por\_genero, color = Genero), hjust = -.05, vjust = 1) +  
 labs(  
 title = "LUCRO MEDIO MUNDIAL DAS DISTRIBUIDORAS POR GENERO",  
 x = "DISTRIBUIDORA",  
 y = "LUCRO MEDIO (MILHPES DE DOLARES)",  
 color = "GENERO"   
 ) +  
 scale\_y\_continuous(breaks = seq(0, 3500, by = 500), label = paste0(seq(0, 3500, by = 500), "MM")) +  
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

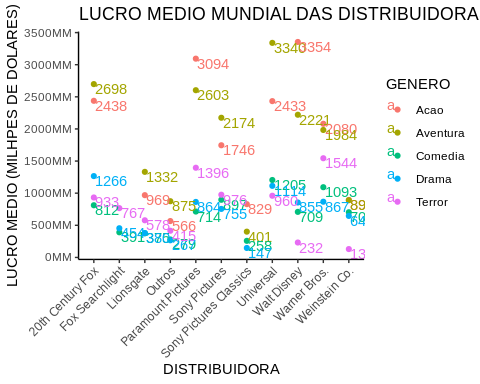


Grafico final

### AULA 3

# EXERCICIO 1