

Exercício Avaliativo - Recursos Computacionais

Parte II (ggplot2)

Antonio Mendes Magalhães Junior

17 de Abril de 2019

Parte II

Objetivo: Ilustrar o uso do pacote `ggplot2` na análise exploratória do banco de dados “Relação de matriculados 2013 a 2017 PRG UFLA”.

Os dados foram cedidos por uma pesquisadora do Departamento de Ciências Humanas (DCH/UFLA), área de sociologia, que conduz o projeto “Mulheres no Ensino Superior”.

O banco de dados refere-se às matrículas nos cursos de graduação da UFLA, segundo ano (2013 a 2017), [area do curso na CAPES, sexo e tipo de ingresso (ampla concorrência e modalidades de cotas)].

1. Estude a proporção de ingressantes (em relação ao ingresso total, i.e., AC + cotas) do sexo feminino nos anos de 2013 e 2017, nos cursos de:
 - Ciências Exatas e engenharias
 - Licenciatura
2. Se comparadas aos homens, as mulheres estão mais representadas em cursos noturnos que em diurnos? Faça o estudo ao longo dos anos.
3. ~~Avalie o número de ingressantes (ou proporção), ao longo dos anos, por área de conhecimento, e segundo sexo e tipo de ingresso (AC e cotas). Represente essas informações em um único gráfico. — Desdobre o gráfico anterior em duas figuras (gráficos), uma para cada tipo de ingresso: um para ampla concorrência (AC) e outra para ingresso por cotas.~~
4. Formule e responda outra pergunta de interesse prático que pode ser respondida nessa análise exploratória.

Essa parte da atividade poderá ser feita durante a próxima aula (11/04), mas aconselho que os dados já estejam organizados e carregados em R.

Prazo máximo de entrega: **13h do dia 17/04/2019**

OBS: O relatório deverá ser feito em Rmarkdown, formato pdf e deverá ser enviado por e-mail (izabela.oliveira@ufla.br).

Primeiramente foram chamadas todas as bibliotecas e os dados utilizados na resolução dos exercícios.

```
library(tinytex)
library(readxl)
library(tidyverse)
dir <- "C:/Users/jrjpm/Desktop/UFLA/Mestrado/Recursos Computacionais/ggplot2/Dados.xlsx"
#"C:/Users/jrjpm/Desktop/Mestrado/Recursos Computacionais/ggplot2/Dados.xlsx";
Dados <- read_excel(dir)
```

Exercício 1

Se comparados os dados do primeiro e do último ano disponíveis (2013 e 2017), é possível observar que houve apenas um leve aumento na proporção de ingressantes do sexo feminino em ambas as áreas estudadas. Entretanto, é interessante verificar a grande diferença de proporção de mulheres entre áreas de Exatas/Engenharias e Licenciatura, já que esta última possui 20% a mais de ingressantes do sexo feminino.

```
Dados <- cbind(Dados[1:10],Dados[19:21])

Dadosf2013Ex <- filter(Dados, ano==2013,
                        areaCAPES=="CIEN_EXATAS" | areaCAPES=="ENGENHARIA");
Dadosf2013L <- filter(Dados, ano==2013, modalidade=="L");

Dadosf2017Ex <- filter(Dados, ano==2017,
                        areaCAPES=="CIEN_EXATAS" | areaCAPES=="ENGENHARIA")
Dadosf2017L <- filter(Dados, ano==2017, modalidade=="L");

#proporção de mulheres (F) em exatas/engenharias em 2013
PFEx13 <- sum(Dadosf2013Ex$C_F+Dadosf2013Ex$AC_F)/sum(Dadosf2013Ex$TOTAL)

#proporção de mulheres (F) em licenciatura em 2013
PFL13 <- sum(Dadosf2013L$C_F+Dadosf2013L$AC_F)/sum(Dadosf2013L$TOTAL)

#proporção de mulheres (F) em exatas/engenharias em 2017
PFEx17 <- sum(Dadosf2017Ex$C_F+Dadosf2017Ex$AC_F)/sum(Dadosf2017Ex$TOTAL)

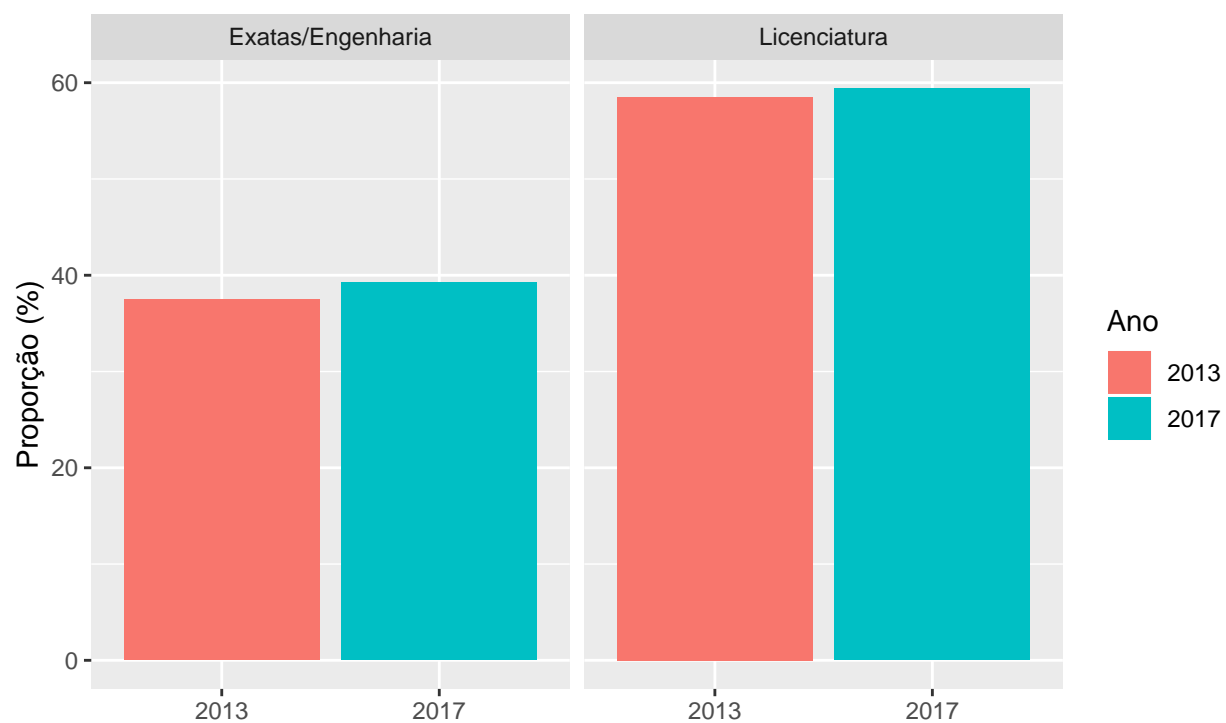
#proporção de mulheres (F) em licenciatura em 2017
PFL17 <- sum(Dadosf2017L$C_F+Dadosf2017L$AC_F)/sum(Dadosf2017L$TOTAL)

dadosprop <- tibble(ano = c("2013", "2013", "2017","2017"),
                    area = c("Exatas/Engenharia", "Licenciatura",
                              "Exatas/Engenharia","Licenciatura"),
                    prop = c(PFEx13 , PFL13, PFEx17,PFL17 ));

ggplot(data = dadosprop) +
  geom_col(mapping = aes(x =ano, y = prop*100,fill=ano))+
  facet_grid(. ~ area)+
  labs(
    title = "Proporção de mulheres ingressantes",
    subtitle = "Comparação entre Exatas/Engenharia e Licenciatura",
    x = " ",
    y = "Proporção (%)",
    fill="Ano")
```

Proporção de mulheres ingressantes

Comparação entre Exatas/Engenharia e Licenciatura



Exercicio 2

O gráfico mostra que, entre 2013 e 2017, com exceção do ano de 2014, o número de ingressantes do sexo feminino sempre é ligeiramente maior no período noturno. Isso provavelmente ocorre em razão da maioria dos cursos da modalidade licenciatura serem em período noturno.

```
DadosN13 <- filter(Dados, turno=="NOTURNO", ano==2013);
DadosD13 <- filter(Dados, turno=="DIURNO", ano==2013);

DadosN14 <- filter(Dados, turno=="NOTURNO", ano==2014);
DadosD14 <- filter(Dados, turno=="DIURNO", ano==2014);

DadosN15 <- filter(Dados, turno=="NOTURNO", ano==2015);
DadosD15 <- filter(Dados, turno=="DIURNO", ano==2015);

DadosN16 <- filter(Dados, turno=="NOTURNO", ano==2016);
DadosD16 <- filter(Dados, turno=="DIURNO", ano==2016);

DadosN17 <- filter(Dados, turno=="NOTURNO", ano==2017);
DadosD17 <- filter(Dados, turno=="DIURNO", ano==2017);

# Proporção de mulheres (F) nos cursos noturnos 13
PF_N13 <- sum(DadosN13$C_F+DadosN13$AC_F)/sum(DadosN13$TOTAL);
# Proporção de mulheres (F) nos cursos diurnos 13
PF_D13 <- sum(DadosD13$C_F+DadosD13$AC_F)/sum(DadosD13$TOTAL);

# Proporção de mulheres (F) nos cursos noturnos 14
PF_N14 <- sum(DadosN14$C_F+DadosN14$AC_F)/sum(DadosN14$TOTAL);
# Proporção de mulheres (F) nos cursos diurnos 14
PF_D14 <- sum(DadosD14$C_F+DadosD14$AC_F)/sum(DadosD14$TOTAL);

# Proporção de mulheres (F) nos cursos noturnos 15
PF_N15 <- sum(DadosN15$C_F+DadosN15$AC_F)/sum(DadosN15$TOTAL);
# Proporção de mulheres (F) nos cursos diurnos 15
PF_D15 <- sum(DadosD15$C_F+DadosD15$AC_F)/sum(DadosD15$TOTAL);

# Proporção de mulheres (F) nos cursos noturnos 16
PF_N16 <- sum(DadosN16$C_F+DadosN16$AC_F)/sum(DadosN16$TOTAL);
# Proporção de mulheres (F) nos cursos diurnos 16
PF_D16 <- sum(DadosD16$C_F+DadosD16$AC_F)/sum(DadosD16$TOTAL);

# Proporção de mulheres (F) nos cursos noturnos 17
PF_N17 <- sum(DadosN17$C_F+DadosN17$AC_F)/sum(DadosN17$TOTAL);
# Proporção de mulheres (F) nos cursos diurnos 17
PF_D17 <- sum(DadosD17$C_F+DadosD17$AC_F)/sum(DadosD17$TOTAL);

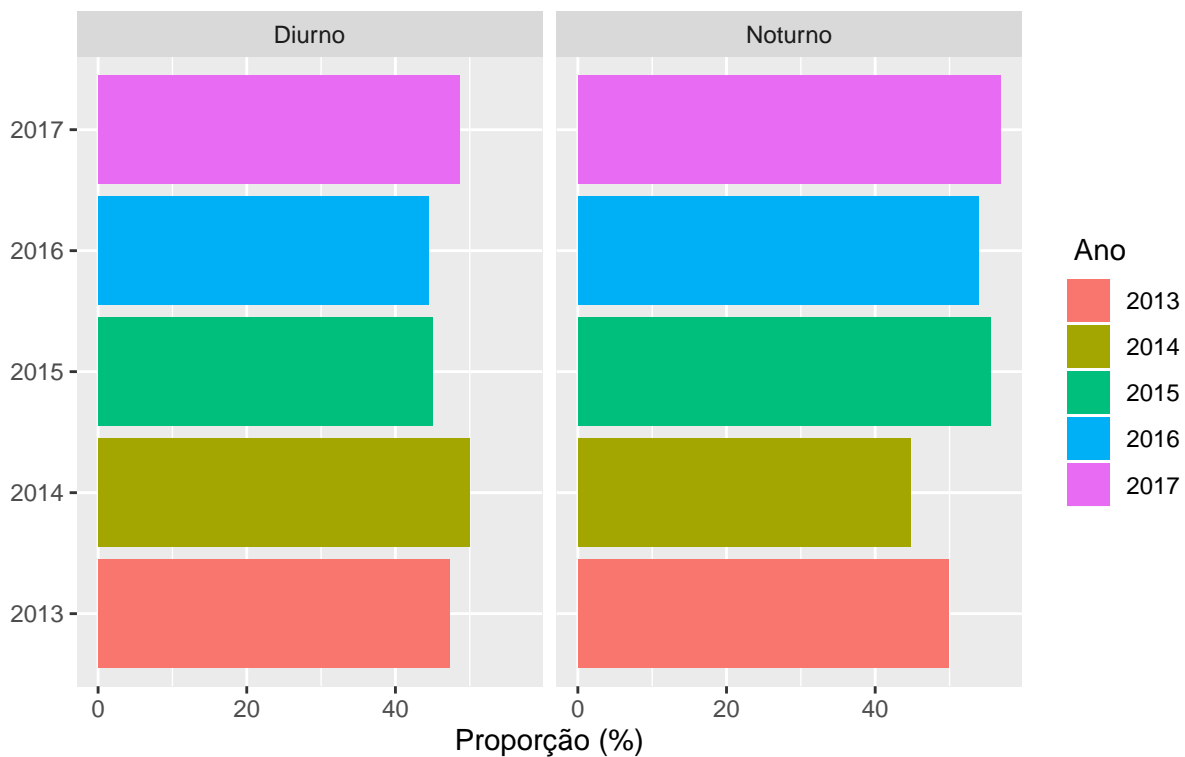
dadosprop <- tibble(ano = c("2013", "2013", "2014", "2014", "2015",
                             "2015", "2016", "2016", "2017", "2017"),
                    turno = c("Noturno", "Diurno", "Noturno", "Diurno", "Noturno",
                              "Diurno", "Noturno", "Diurno", "Noturno", "Diurno"),
                    prop = c(PF_N13, PF_D13, PF_N14, PF_D14, PF_N15,
                              PF_D15, PF_N16, PF_D16, PF_N17, PF_D17));

ggplot(data = dadosprop) +
```

```
geom_col(mapping = aes(x =ano, y = prop*100,fill=ano))+
facet_grid(. ~ turno)+
coord_flip()+
labs(
  title = "Proporção de mulheres ingressantes",
  subtitle = "Análise por período ao longo dos anos",
  x = " ",
  y = "Proporção (%)",
  fill = " Ano");
```

Proporção de mulheres ingressantes

Análise por período ao longo dos anos



Exercício 4

Visando a comprovação de que no curso de nutrição há uma maior proporção de ingressantes do sexo feminino do que no curso de Engenharia de Controle e Automação, foi plotado um gráfico comparativo. O gráfico mostra uma grande disparidade nas proporções de mulheres entre os referidos cursos ao longo de todos os anos estudados. O curso de Nutrição apresenta um percentual de mulheres ingressantes próximo a 85%, enquanto que o curso de Engenharia de Controle e Automação apresenta um percentual próximo a 20%.

```
# cursos estudados (indique dois)
CompC <- c("ENG_CONT", "NUTRICA0");

Dados13 <- filter(Dados, curso==CompC[1], ano==2013)
Dados14 <- filter(Dados, curso==CompC[1], ano==2014)
Dados15 <- filter(Dados, curso==CompC[1], ano==2015)
Dados16 <- filter(Dados, curso==CompC[1], ano==2016)
Dados17 <- filter(Dados, curso==CompC[1], ano==2017)

PF_13_C1 <- sum(Dados13$C_F+Dados13$AC_F)/sum(Dados13$TOTAL);
PF_14_C1 <- sum(Dados14$C_F+Dados14$AC_F)/sum(Dados14$TOTAL);
PF_15_C1 <- sum(Dados15$C_F+Dados15$AC_F)/sum(Dados15$TOTAL);
PF_16_C1 <- sum(Dados16$C_F+Dados16$AC_F)/sum(Dados16$TOTAL);
PF_17_C1 <- sum(Dados17$C_F+Dados17$AC_F)/sum(Dados17$TOTAL);

Dados13 <- filter(Dados, curso==CompC[2], ano==2013)
Dados14 <- filter(Dados, curso==CompC[2], ano==2014)
Dados15 <- filter(Dados, curso==CompC[2], ano==2015)
Dados16 <- filter(Dados, curso==CompC[2], ano==2016)
Dados17 <- filter(Dados, curso==CompC[2], ano==2017)

PF_13_C2 <- sum(Dados13$C_F+Dados13$AC_F)/sum(Dados13$TOTAL);
PF_14_C2 <- sum(Dados14$C_F+Dados14$AC_F)/sum(Dados14$TOTAL);
PF_15_C2 <- sum(Dados15$C_F+Dados15$AC_F)/sum(Dados15$TOTAL);
PF_16_C2 <- sum(Dados16$C_F+Dados16$AC_F)/sum(Dados16$TOTAL);
PF_17_C2 <- sum(Dados17$C_F+Dados17$AC_F)/sum(Dados17$TOTAL);

dadosprop <- tibble(ano = c("2013", "2014", "2015", "2016", "2017",
                             "2013", "2014", "2015", "2016", "2017"),
                    curso = c(CompC[1], CompC[1], CompC[1], CompC[1], CompC[1],
                              CompC[2], CompC[2], CompC[2], CompC[2], CompC[2]),
                    prop = c(PF_13_C1, PF_14_C1, PF_15_C1, PF_16_C1, PF_17_C1,
                              PF_13_C2, PF_14_C2, PF_15_C2, PF_16_C2, PF_17_C2));

ggplot(data = dadosprop) +
  geom_col(mapping = aes(x = ano, y = prop*100, fill=ano)) +
  coord_flip() +
  facet_grid(. ~ curso) +
  labs(
    title = paste0("Proporção de mulheres ingressantes \nnos cursos de: ", CompC[1],
                  " vs ", CompC[2]),
    subtitle = "Análise ao longo dos anos",
    x = " ",
    y = "Proporção (%)",
    fill="Ano")
```

Proporção de mulheres ingressantes nos cursos de: ENG_CONT vs NUTRICA0

Análise ao longo dos anos

