

Exercício Avaliativo - Recursos Computacionais

Atividade 2 - dplyr

Antonio Mendes Magalhães Junior

26 de Abril de 2019

Atividade II - Uso do pacote dplyr()

Ilustre o uso das cinco funções básicas do pacote `dplyr` (`filter()`, `arrange()`, `select()`, `mutate()` e `summarize()`) com o banco de dados de ingressantes na UFLA (período de 2013 a 2017).

Exemplifique cada função pelo menos 2 vezes e utilize o comando `group_by` e o operador `%>%` (*pipe*) ao menos uma vez.

Prazo de entrega: Dia 26/04/2019

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

Referência

WICKHAM, H.; GROLEMUND, G. (2016) *R for Data Science*. Sebastopol: O'Reilly.

Importando os dados

- Primeiramente foram importadas todas as bibliotecas e os dados utilizados na atividade.

```
library(tinytex)
library(readxl)
library(tidyverse)
dir <- "C:/Users/jrjpm/Desktop/Mestrado/Recursos Computacionais/dplyr/Dados.xlsx";
Dados <- read_excel(dir)
Dados <- cbind(Dados[1:10],Dados[19:21])
head(Dados)
```

	ano	curso	modalidade	turno	areaCAPES	codigo	C_M	C_F	AC_M	AC_F
1	2013	AGRONOMIA		B DIURNO	CIEN_AGRA	G001	22	8	109	59
2	2013	ZOOTECNIA		B DIURNO	CIEN_AGRA	G002	5	11	45	40
3	2013	ENG_AGRIC		B DIURNO	CIEN_AGRA	G003	8	1	34	7
4	2013	ENG_FLORES		B DIURNO	CIEN_AGRA	G005	10	7	48	36
5	2013	MED_VETER		B DIURNO	CIEN_AGRA	G007	5	11	33	51
6	2013	ADMINISTRA		B DIURNO	CIE_SOCAPL	G009	5	5	41	28
	TOTAL	TOTAL_C	TOTAL_AC							
1	198	30	168							
2	101	16	85							
3	50	9	41							
4	101	17	84							
5	100	16	84							
6	79	10	69							

Função *filter*

- Filtrando os cursos da área de Ciências Agrárias no ano de 2017.

```
n_dados <- filter(Dados, areaCAPES=="CIEN_AGRA", ano==2017)
head(n_dados)
```

	ano	curso	modalidade	turno	areaCAPES	codigo	C_M	C_F	AC_M	AC_F
1	2017	AGRONOMIA	B	DIURNO	CIEN_AGRA	G001	50	36	74	22
2	2017	ZOOTECNIA	B	DIURNO	CIEN_AGRA	G002	17	31	20	25
3	2017	ENG_AGRIC	B	DIURNO	CIEN_AGRA	G003	18	6	20	2
4	2017	ENG_FLORES	B	DIURNO	CIEN_AGRA	G005	23	21	21	28
5	2017	MED_VETER	B	DIURNO	CIEN_AGRA	G007	16	29	10	34
6	2017	ENG_ALIMEN	B	DIURNO	CIEN_AGRA	G011	16	29	14	29
		TOTAL	TOTAL_C	TOTAL_AC						
1	182	86	96							
2	93	48	45							
3	46	24	22							
4	93	44	49							
5	89	45	44							
6	88	45	43							

- Filtrando os cursos de Engenharia de Controle e Automação e Nutrição no ano de 2015.

```
n_dados <- filter(Dados, curso==c("ENG_CONT", "NUTRICA0"), ano==2015)
head(n_dados)
```

	ano	curso	modalidade	turno	areaCAPES	codigo	C_M	C_F	AC_M	AC_F
1	2015	ENG_CONT	B	DIURNO	ENGENHARIA	G022	29	9	44	13
2	2015	NUTRICA0	B	DIURNO	CIEN_SAUDE	G023	6	30	11	42
		TOTAL	TOTAL_C	TOTAL_AC						
1	95	38	57							
2	89	36	53							

Função *arrange*

- Ordenando os dados pelo turno dos cursos.

```
n_dados <- arrange(Dados,turno)
head(n_dados)
```

	ano	curso	modalidade	turno	areaCAPES	codigo	C_M	C_F	AC_M	AC_F
1	2013	AGRONOMIA	B	DIURNO	CIEN_AGRA	G001	22	8	109	59
2	2013	ZOOTECNIA	B	DIURNO	CIEN_AGRA	G002	5	11	45	40
3	2013	ENG_AGRIC	B	DIURNO	CIEN_AGRA	G003	8	1	34	7
4	2013	ENG_FLORES	B	DIURNO	CIEN_AGRA	G005	10	7	48	36
5	2013	MED_VETER	B	DIURNO	CIEN_AGRA	G007	5	11	33	51
6	2013	ADMINISTRA	B	DIURNO	CIE_SOCAPL	G009	5	5	41	28
	TOTAL	TOTAL_C	TOTAL_AC							
1	198	30	168							
2	101	16	85							
3	50	9	41							
4	101	17	84							
5	100	16	84							
6	79	10	69							

- Ordenando os dados, de forma decrescente, por total de mulheres ingressantes por cotas (C_F) em 2017 na modalidade licenciatura.

```
n_dados <- arrange(Dados,desc(C_F))
n_dados <- filter( n_dados ,modalidade=="L" & ano==2017)
head(n_dados)
```

	ano	curso	modalidade	turno	areaCAPES	codigo	C_M	C_F	AC_M	AC_F
1	2017	PEDAGOGIA	L	NOTURNO	CIEN_HUMA	G036	5	44	11	39
2	2017	LETRAS	L	NOTURNO	LING_LET_AR	G025	9	40	11	30
3	2017	QUIMICA	L	DIURNO	CIEN_EXATAS	G013	16	35	20	26
4	2017	CIE_BIOLLIC	L	DIURNO	CIEN_EXATAS	G020	12	14	7	17
5	2017	EDU_FISLIC	L	DIURNO	CIEN_SAUDE	G028	18	12	11	12
6	2017	MATEMAT	L	NOTURNO	CIEN_EXATAS	G015	22	8	14	12
	TOTAL	TOTAL_C	TOTAL_AC							
1	99	49	50							
2	90	49	41							
3	97	51	46							
4	50	26	24							
5	53	30	23							
6	56	30	26							

Função *select*

- Organizando os dados retirando as colunas de modalidade, área CAPES e código dos cursos.

```
n_dados <- select(Dados, -c(modalidade, codigo, areaCAPES))  
head(n_dados)
```

	ano	curso	turno	C_M	C_F	AC_M	AC_F	TOTAL	TOTAL_C	TOTAL_AC
1	2013	AGRONOMIA	DIURNO	22	8	109	59	198	30	168
2	2013	ZOOTECNIA	DIURNO	5	11	45	40	101	16	85
3	2013	ENG_AGRIC	DIURNO	8	1	34	7	50	9	41
4	2013	ENG_FLORES	DIURNO	10	7	48	36	101	17	84
5	2013	MED_VETER	DIURNO	5	11	33	51	100	16	84
6	2013	ADMINISTRA	DIURNO	5	5	41	28	79	10	69

- Organizando os dados deixando apenas as colunas de ano, curso e total de ingressantes mulheres por cotas e ampla concorrência. Após isso, ordenando os dados, de forma decrescente, pelo total de ingressantes do sexo feminino por cotas.

```
n_dados <- select(Dados, c(ano, curso, C_F, AC_F))  
n_dados <- arrange(n_dados, desc(C_F))  
head(n_dados)
```

	ano	curso	C_F	AC_F
1	2016	ABI	68	70
2	2015	ABI	63	86
3	2017	ABI	61	73
4	2017	PEDAGOGIA	44	39
5	2016	NUTRICAO	42	40
6	2016	PEDAGOGIA	41	44

Função *mutate*

- Adicionando uma coluna de proporção de ingressantes do sexo feminino por cotas e depois ordenando os dados, de forma decrescente, por esta proporção, no ano de 2017.

```
n_dados <- mutate(Dados,prop=100*C_F/TOTAL)
n_dados <- filter(n_dados,ano==2017)
n_dados <- select(n_dados, ano, curso, modalidade, prop)
n_dados <- arrange(n_dados,desc(prop))
head(n_dados)
```

	ano	curso	modalidade	prop
1	2017	LETRAS	L	44.44444
2	2017	PEDAGOGIA	L	44.44444
3	2017	NUTRICA0	B	42.85714
4	2017	QUIMICA	L	36.08247
5	2017	CIE_BIOLBAC	B	34.14634
6	2017	ZOOTECNIA	B	33.33333

- Usando *mutate* para adicionar uma coluna de proporções de homens e uma de proporções de mulheres. Em seguida, usando a função *filter* para filtrar os dados apenas pelo ano de 2017 e depois utilizando *select* para utilizar apenas as colunas indicadas. Por fim, utilizando a função *arrange* para ordenar os dados de forma decrescente e proporcional em relação ao número de mulheres em cada curso.

```
n_dados <- mutate(Dados, proph=(100*(C_M+AC_M)/TOTAL), propm=100-proph)
n_dados <- filter(n_dados,ano==2017)
n_dados <- select(n_dados, ano, curso, modalidade, proph, propm)
n_dados <- arrange(n_dados,desc(propm))
head(n_dados)
```

	ano	curso	modalidade	proph	propm
1	2017	PEDAGOGIA	L	16.16162	83.83838
2	2017	NUTRICA0	B	18.68132	81.31868
3	2017	LETRAS	L	22.22222	77.77778
4	2017	MED_VETER	B	29.21348	70.78652
5	2017	DIREITO	B	29.24528	70.75472
6	2017	ENG_ALIMEN	B	34.09091	65.90909

- Repetindo o código anterior utilizando o operador *pipe* (*%>%*):

```
n_dados <- Dados %>%
mutate(proph=(100*(C_M+AC_M)/TOTAL), propm=100-proph) %>%
filter(ano==2017) %>%
select(ano, curso, modalidade, proph, propm) %>%
arrange(desc(propm))
head(n_dados)
```

	ano	curso	modalidade	proph	propm
1	2017	PEDAGOGIA	L	16.16162	83.83838
2	2017	NUTRICA0	B	18.68132	81.31868
3	2017	LETRAS	L	22.22222	77.77778
4	2017	MED_VETER	B	29.21348	70.78652
5	2017	DIREITO	B	29.24528	70.75472

6 2017 ENG_ALIMEN

B 34.09091 65.90909

Função *summarise*

- Fazendo a média de ingressantes do sexo femininos por cotas, entre os anos de 2013 e 2017, para cada um dos cursos.

```
n_dados<- Dados %>%  
group_by(curso) %>%  
summarise(mean(C_F))  
head(n_dados)
```

```
# A tibble: 6 x 2  
  curso      `mean(C_F)`  
  <chr>      <dbl>  
1 ABI          54.2  
2 ADM_PUBLIC    19  
3 ADMINISTRA    11.8  
4 AGRONOMIA     23.8  
5 CIE_BIOLBAC   11.4  
6 CIE_BIOLLIC   11.6
```

- Fazendo a média da proporção dos ingressantes de ambos os sexos por cotas, entre todos os cursos, para cada turno e ano.

```
Dados %>%  
group_by(ano, turno) %>%  
summarise(mean((C_F+C_M)/TOTAL))
```

```
# A tibble: 10 x 3  
# Groups:   ano [5]  
  ano turno      `mean((C_F + C_M)/TOTAL)`  
  <dbl> <chr>      <dbl>  
1  2013 DIURNO          0.157  
2  2013 NOTURNO         0.179  
3  2014 DIURNO          0.273  
4  2014 NOTURNO         0.288  
5  2015 DIURNO          0.392  
6  2015 NOTURNO         0.398  
7  2016 DIURNO          0.494  
8  2016 NOTURNO         0.479  
9  2017 DIURNO          0.521  
10 2017 NOTURNO         0.501
```