

# Manipulação e Apresentação de Dados

Guilherme Bovi Ambrosano



▶ broom

▶ purrr

▶ ggplot2

▶ Exercícios

▶ Exemplos

▶ Exercícios

broom



```
modelo.aov <- aov(Sepal.Length ~ Species, data=iris)
modelo.lm  <- lm(Sepal.Width ~ Petal.Width, data=iris)
modelo.wt  <- wilcox.test(iris$Petal.Length, g=iris$Species)
```



```
library(broom)
tidy(modelo.aov)
```

```
## # A tibble: 2 x 6
##   term      df sumsq meansq statistic  p.value
##   <chr>    <dbl> <dbl>  <dbl>    <dbl>    <dbl>
## 1 Species      2  63.2  31.6      119.  1.67e-31
## 2 Residuals   147  39.0   0.265      NA    NA
```

```
tidy(modelo.lm)
```

```
## # A tibble: 2 x 5
##   term      estimate std.error statistic  p.value
##   <chr>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)  3.31    0.0621    53.3  1.84e-98
## 2 Petal.Width -0.209   0.0437    -4.79  4.07e- 6
```

```
tidy(modelo.wt)
```

```
## # A tibble: 1 x 4
##   statistic p.value method      alternat
##   <dbl>    <dbl> <chr>    <chr>
## 1    11325 2.25e-26 Wilcoxon signed rank test with continuity corr~ two.sided
```



```
glance(modelo.aov)
```

```
## # A tibble: 1 x 6
##   logLik   AIC   BIC deviance  nobs r.squared
##   <dbl> <dbl> <dbl>     <dbl> <int>     <dbl>
## 1  -112.  231.  243.     39.0   150     0.619
```

```
glance(modelo.lm)
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic    p.value    df logLik   AIC   BI
##   <dbl>         <dbl> <dbl>     <dbl>     <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    0.134         0.128 0.407     22.9 0.00000407     1  -77.0  160.  169
## # i 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
glance(modelo.wt)
```

```
## # A tibble: 1 x 4
##   statistic p.value method                                alterna
##   <dbl>     <dbl> <chr>                                <chr>
## 1    11325 2.25e-26 Wilcoxon signed rank test with continuity corr~ two.sid
```

```
augment(modelo.lm, interval="confidence")
```

```
## # A tibble: 150 x 10
##   Sepal.Width Petal.Width .fitted .lower .upper .resid .hat .sig
##   <dbl>         <dbl>   <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>
## 1         3.5         0.2    3.27  3.16  3.38  0.233  0.0182  0.4
## 2          3         0.2    3.27  3.16  3.38 -0.267  0.0182  0.4
## 3         3.2         0.2    3.27  3.16  3.38 -0.0666  0.0182  0.4
## 4         3.1         0.2    3.27  3.16  3.38 -0.167  0.0182  0.4
## 5         3.6         0.2    3.27  3.16  3.38  0.333  0.0182  0.4
## 6         3.9         0.4    3.22  3.13  3.32  0.675  0.0140  0.4
## 7         3.4         0.3    3.25  3.14  3.35  0.154  0.0160  0.4
## 8         3.4         0.2    3.27  3.16  3.38  0.133  0.0182  0.4
## 9         2.9         0.2    3.27  3.16  3.38 -0.367  0.0182  0.4
## 10        3.1         0.1    3.29  3.17  3.40 -0.187  0.0206  0.4
## # i 140 more rows
## # i 1 more variable: .std.resid <dbl>
```

purrr





```
lapply(seq(1,5), function(x) {  
  paste(x, "é", ifelse(x%%2==0, "par", "ímpar"))  
})
```

```
## [[1]]  
## [1] "1 é ímpar"  
##  
## [[2]]  
## [1] "2 é par"  
##  
## [[3]]  
## [1] "3 é ímpar"  
##  
## [[4]]  
## [1] "4 é par"  
##  
## [[5]]  
## [1] "5 é ímpar"
```



```
library(purrr)
map(seq(1,5), function(x) {
  paste(x, "é", ifelse(x%%2==0, "par", "ímpar"))
})
```

```
## [[1]]
## [1] "1 é ímpar"
##
## [[2]]
## [1] "2 é par"
##
## [[3]]
## [1] "3 é ímpar"
##
## [[4]]
## [1] "4 é par"
##
## [[5]]
## [1] "5 é ímpar"
```



```
map_dfr(seq(1,5), function(x) {  
  tibble(par=paste(x, "é", ifelse(x%%2==0, "par", "ímpar")))  
})
```

```
## # A tibble: 5 x 1  
##   par  
##   <chr>  
## 1 1 é ímpar  
## 2 2 é par  
## 3 3 é ímpar  
## 4 4 é par  
## 5 5 é ímpar
```



```
map_dfc(tibble(um=1, dois=2, tres=3), function(x) {  
  x * seq(1,5)  
})
```

```
## # A tibble: 5 x 3  
##       um  dois  tres  
##   <dbl> <dbl> <dbl>  
## 1     1     2     3  
## 2     2     4     6  
## 3     3     6     9  
## 4     4     8    12  
## 5     5    10    15
```

```
library(agridat)
caribbean.maize
```

##		isle	site	block	plot	trt	ears	yield
## 1		Antigua	DBAN	B1	1	T111	42	4.96
## 2		Antigua	DBAN	B1	2	T000	41	3.94
## 3		Antigua	DBAN	B1	3	T311	49	6.35
## 4		Antigua	DBAN	B1	4	T202	48	5.56
## 5		Antigua	DBAN	B1	5	T111	45	5.36
## 6		Antigua	DBAN	B1	6	T220	46	6.18
## 7		Antigua	DBAN	B1	7	T113	42	4.71
## 8		Antigua	DBAN	B1	8	T131	44	6.03
## 9		Antigua	DBAN	B1	9	T022	42	2.88
## 10		Antigua	DBAN	B2	10	T222	44	5.68
## 11		Antigua	DBAN	B2	11	T311	42	5.80
## 12		Antigua	DBAN	B2	12	T020	42	4.16
## 13		Antigua	DBAN	B2	13	T200	46	4.90
## 14		Antigua	DBAN	B2	14	T111	44	5.25
## 15		Antigua	DBAN	B2	15	T131	48	5.80
## 16		Antigua	DBAN	B2	16	T002	46	2.18



```
caribbean.maize %>%  
  group_split(isle)
```

```
## <list_of<  
##   tbl_df<  
##     isle : factor<2a195>  
##     site : factor<459f2>  
##     block: factor<cfca5>  
##     plot : integer  
##     trt   : factor<ed0b9>  
##     ears  : integer  
##     yield: double  
##   >  
## >[2]>  
## [[1]]  
## # A tibble: 288 x 7  
##   isle   site block plot trt   ears yield  
##   <fct> <fct> <fct> <int> <fct> <int> <dbl>  
## 1 Antigua DBAN  B1      1 T111    42  4.96  
## 2 Antigua DBAN  B1      2 T000    41  3.04
```

```
caribbean.maize %>%
  group_split(isle) %>%
  map(function(x) x %>%
    pivot_wider(id_cols=c(block,plot), names_from=site,
                 values_from=yield))
```

```
## [[1]]
```

```
## # A tibble: 36 x 10
```

```
##   block plot  DBAN  LFAN  TEAN  WEAN  WLAN  NSAN  OVAN  ORA
```

```
##   <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1 B1      1  4.96  2.92  1.27  4.02  2      2.43  2.04  5.2
```

```
## 2 B1      2  3.94  1.68  2.1   5.8   2.39  1.28  3.88  6.8
```

```
## 3 B1      3  6.35  4.5   2.37  2.16  2.04  0.83  5.53  7.7
```

```
## 4 B1      4  5.56  4.74  2.08  5.31  1.64  2.44  6.61  6.7
```

```
## 5 B1      5  5.36  4.44  1.74  5.12  1.83  1.3   2.97  6.5
```

```
## 6 B1      6  6.18  4.66  3.16  5.98  2.08  1.34  4.47  6.4
```

```
## 7 B1      7  4.71  2.94  4.55  5.46  3.13  2.06  4.06  6.7
```

```
## 8 B1      8  6.03  1.6   3.2   3.45  1.96  3.02  5.02  6.7
```

```
## 9 B1      9  2.88  1.82  3.69  4.96  1.3   2.18  4.56  6.9
```

```
## 10 B2     10  5.68  4.44  1.52  4.26  2.16  2.25  3.5   2.9
```



```
caribbean.maize %>%
  group_split(isle) %>%
  map(\(x) x %>%
    pivot_wider(id_cols=c(block,plot), names_from=site,
                 values_from=yield))
```

```
## [[1]]
```

```
## # A tibble: 36 x 10
```

##	block	plot	DBAN	LFAN	TEAN	WEAN	WLAN	NSAN	OVAN	ORAN
##	<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
##	1 B1	1	4.96	2.92	1.27	4.02	2	2.43	2.04	5.2
##	2 B1	2	3.94	1.68	2.1	5.8	2.39	1.28	3.88	6.8
##	3 B1	3	6.35	4.5	2.37	2.16	2.04	0.83	5.53	7.7
##	4 B1	4	5.56	4.74	2.08	5.31	1.64	2.44	6.61	6.7
##	5 B1	5	5.36	4.44	1.74	5.12	1.83	1.3	2.97	6.5
##	6 B1	6	6.18	4.66	3.16	5.98	2.08	1.34	4.47	6.4
##	7 B1	7	4.71	2.94	4.55	5.46	3.13	2.06	4.06	6.7
##	8 B1	8	6.03	1.6	3.2	3.45	1.96	3.02	5.02	6.7
##	9 B1	9	2.88	1.82	3.69	4.96	1.3	2.18	4.56	6.9
##	10 B2	10	5.68	4.44	1.52	4.26	2.16	2.25	3.5	2.9





```
caribbean.maize %>%
  group_split(isle) %>%
  map(. %>%
    pivot_wider(id_cols=c(block,plot), names_from=site,
                 values_from=yield))
```

```
## [[1]]
```

```
## # A tibble: 36 x 10
```

```
##   block plot DBAN  LFAN  TEAN  WEAN  WLAN  NSAN  OVAN  ORA
```

```
##   <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1 B1      1  4.96  2.92  1.27  4.02  2      2.43  2.04  5.2
```

```
## 2 B1      2  3.94  1.68  2.1   5.8   2.39  1.28  3.88  6.8
```

```
## 3 B1      3  6.35  4.5   2.37  2.16  2.04  0.83  5.53  7.7
```

```
## 4 B1      4  5.56  4.74  2.08  5.31  1.64  2.44  6.61  6.7
```

```
## 5 B1      5  5.36  4.44  1.74  5.12  1.83  1.3   2.97  6.5
```

```
## 6 B1      6  6.18  4.66  3.16  5.98  2.08  1.34  4.47  6.4
```

```
## 7 B1      7  4.71  2.94  4.55  5.46  3.13  2.06  4.06  6.7
```

```
## 8 B1      8  6.03  1.6   3.2   3.45  1.96  3.02  5.02  6.7
```

```
## 9 B1      9  2.88  1.82  3.69  4.96  1.3   2.18  4.56  6.9
```

```
## 10 B2     10  5.68  4.44  1.52  4.26  2.16  2.25  3.5   2.9
```



```
iris %>%  
  select(-Species) %>%  
  as.list() %>%  
  map(\(x) x %>%  
    enframe() %>%  
    cbind(Species=iris$Species) %>%  
    aov(value ~ Species, data=.) %>%  
    residuals() %>%  
    shapiro.test() %>%  
    tidy()) %>%  
  bind_rows(.id="Variável")
```

```
## # A tibble: 4 x 4  
##   Variável      statistic p.value method  
##   <chr>         <dbl>   <dbl> <chr>  
## 1 Sepal.Length 0.988 0.219 Shapiro-Wilk normality test  
## 2 Sepal.Width 0.989 0.323 Shapiro-Wilk normality test  
## 3 Petal.Length 0.981 0.0368 Shapiro-Wilk normality test  
## 4 Petal.Width 0.972 0.00387 Shapiro-Wilk normality test
```



```
iris %>%  
  select(starts_with("Sepal")) %>%  
  as.list() %>%  
  map(. %>%  
    enframe() %>%  
    cbind(Species=iris$Species) %>%  
    aov(value ~ Species, data=.) %>%  
    tidy()) %>%  
  bind_rows(.id="Variável")
```

```
## # A tibble: 4 x 7
```

##	Variável	term	df	sumsq	meansq	statistic	p.val
##	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	Sepal.Length	Species	2	63.2	31.6	119.	1.67e-
## 2	Sepal.Length	Residuals	147	39.0	0.265	NA	NA
## 3	Sepal.Width	Species	2	11.3	5.67	49.2	4.49e-
## 4	Sepal.Width	Residuals	147	17.0	0.115	NA	NA



```
iris %>%  
  select(starts_with("Sepal")) %>%  
  as.list() %>%  
  map_dfr(. %>%  
    enframe() %>%  
    cbind(Species=iris$Species) %>%  
    aov(value ~ Species, data=.) %>%  
    tidy())
```

```
## # A tibble: 4 x 6  
##   term          df sumsq meansq statistic    p.value  
##   <chr>      <dbl> <dbl>   <dbl>      <dbl>    <dbl>  
## 1 Species         2  63.2  31.6      119.  1.67e-31  
## 2 Residuals     147  39.0   0.265      NA     NA  
## 3 Species         2  11.3   5.67      49.2  4.49e-17  
## 4 Residuals     147  17.0   0.115      NA     NA
```

```
cor(iris[-5])
```

```
##           Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length      1.0000000 -0.1175698    0.8717538    0.8179411
## Sepal.Width       -0.1175698    1.0000000   -0.4284401   -0.3661259
## Petal.Length      0.8717538  -0.4284401    1.0000000    0.9628654
## Petal.Width       0.8179411  -0.3661259    0.9628654    1.0000000
```

```
cor.test(iris[-5])
```

```
## Error in cor.test.default(iris[-5]): 'X' deve ser um vetor nu
```



```
expand_grid(Var1=names(iris)[-5],  
            Var2=names(iris)[-5]) %>%  
  pmap_dbl(\(Var1, Var2){  
    cor.test(iris[,Var1], iris[,Var2])$p.value  
  }) %>%  
  matrix(nrow=4)
```

```
##           [,1]           [,2]           [,3]           [,4]  
## [1,] 0.000000e+00 1.518983e-01 1.038667e-47 2.325498e-37  
## [2,] 1.518983e-01 0.000000e+00 4.513314e-08 4.073229e-06  
## [3,] 1.038667e-47 4.513314e-08 0.000000e+00 4.675004e-86  
## [4,] 2.325498e-37 4.073229e-06 4.675004e-86 0.000000e+00
```



```
expand_grid(Var1=names(iris)[-5],  
            Var2=names(iris)[-5]) %>%  
  pmap_chr(\(Var1, Var2){  
    if(Var1==Var2) return("-")  
    cor.test(iris[,Var1], iris[,Var2]) %>%  
      .$p.value %>%  
      scales::pvalue(acc=.0001, dec=","")  
  }) %>%  
  matrix(nrow=4)
```

```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "-"      "0,1519"  "<0,0001"  "<0,0001"  
## [2,] "0,1519"  "-"      "<0,0001"  "<0,0001"  
## [3,] "<0,0001"  "<0,0001"  "-"      "<0,0001"  
## [4,] "<0,0001"  "<0,0001"  "<0,0001"  "-"
```



```
map2(c(1,2,3,4), c("a","b","c","d"),  
      \(num, chr) paste0("Número: ", num, "; Caractere: ", chr))
```

```
## [[1]]  
## [1] "Número: 1; Caractere: a"  
##  
## [[2]]  
## [1] "Número: 2; Caractere: b"  
##  
## [[3]]  
## [1] "Número: 3; Caractere: c"  
##  
## [[4]]  
## [1] "Número: 4; Caractere: d"
```



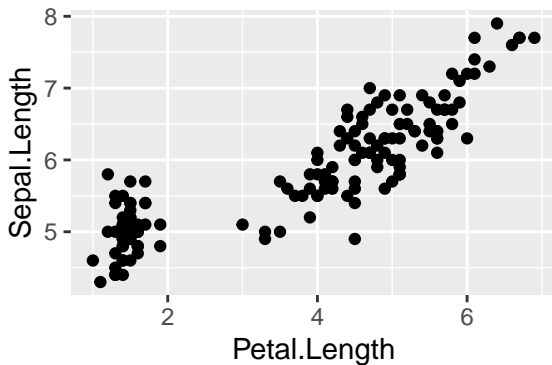
ggplot2



```
library(ggplot2)
iris %>%
  ggplot(aes(x=Species, y=Sepal.Length)) +
  geom_boxplot()
```

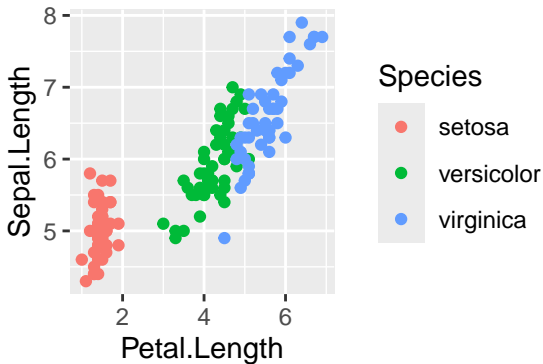


```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length)) +  
  geom_point()
```



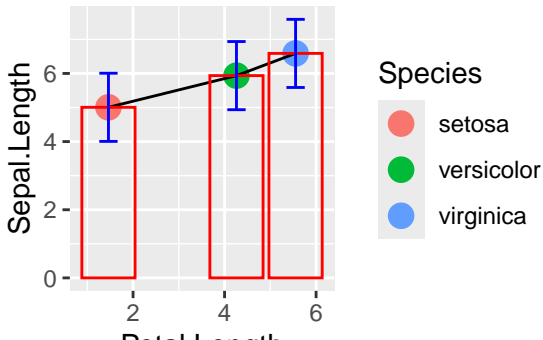


```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length,  
             color=Species)) +  
  geom_point()
```





```
iris %>%  
  group_by(Species) %>%  
  summarise_at(vars(Petal.Length, Sepal.Length), mean) %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length, col=Species)) +  
  geom_point(size=4) +  
  geom_line(col="black") +  
  geom_col(col="red", fill="transparent") +  
  geom_errorbar(aes(ymax=Sepal.Length+1, ymin=Sepal.Length-1),  
                col="blue", width=.4)
```





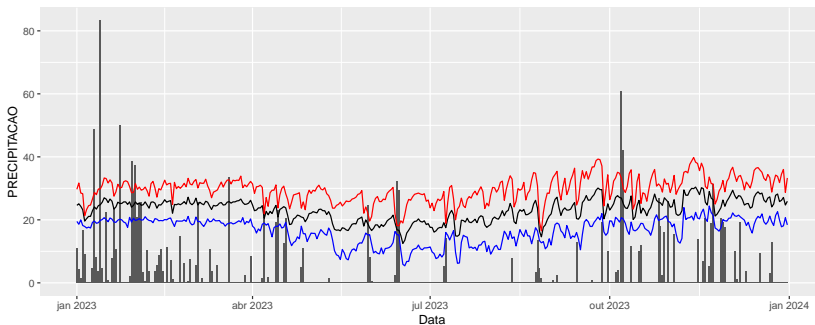
```
linhas <- read_lines(
  "http://www.leb.esalq.usp.br/leb/exceldados/DCE2023.TXT")

inicio <- which(str_starts(linhas, "="))[c(1, seq(3, 37, 3))]
final <- which(str_starts(linhas, "="))[c(2, seq(5,37,3), 37)]
pular <- sapply(seq(1,13), function(i){
  seq(inicio[i], final[i])
})
pular2 <- which(linhas=="")

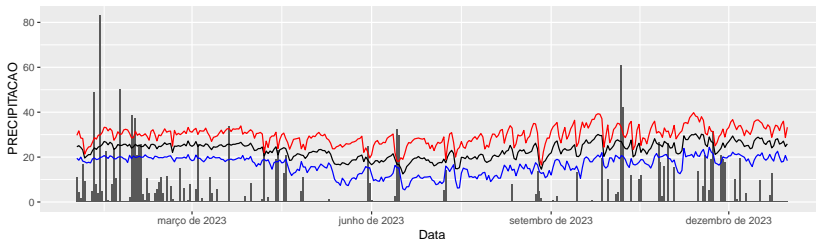
linhas2 <- linhas[-c(unlist(pular),pular2)]

dados_meteorologicos <- as_tibble(linhas2) %>%
  separate(value,
    c("No", "ANO", "DIA", "MES", "R.GLOBA",
      "INSOLACAO", "PRECIPITACAO", "UMIDADE RELATIV",
      "VENTO MAXIMO", "VENTO MEDIO", "TEMPER MAXIMA",
      "TEMPER MINIMA", "TEMPER MEDIA", "EVAPORACAO"),
    sep=" +") %>%
  unite(Data, DIA, MES, ANO) %>%
  mutate(Data=lubridate::dmy(Data)) %>%
  mutate_at(vars(-Data), str_replace, ",", ".") %>%
  mutate_at(vars(-Data), parse_number) %>%
  filter(!is.na(Data))
```

```
dados_meteorologicos %>%
  ggplot(aes(x=Data)) +
  geom_col(aes(y=PRECIPITACAO)) +
  geom_line(aes(y=`TEMPER MEDIA`)) +
  geom_line(aes(y=`TEMPER MINIMA`, col="blue")) +
  geom_line(aes(y=`TEMPER MAXIMA`, col="red"))
```

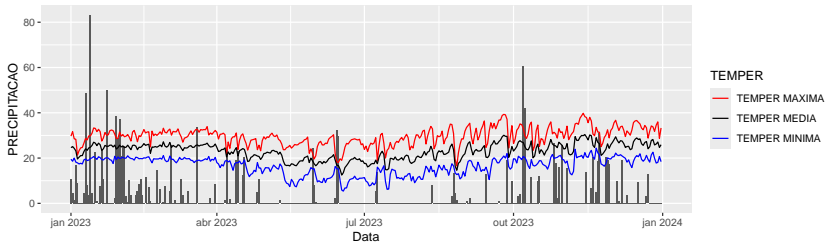


```
dados_meteorologicos %>%
  ggplot(aes(x=Data)) +
  scale_x_date(date_breaks="3 months",
               date_labels="%B de %Y") +
  geom_col(aes(y=PRECIPITACAO)) +
  geom_line(aes(y=`TEMPER MEDIA`)) +
  geom_line(aes(y=`TEMPER MINIMA`), col="blue") +
  geom_line(aes(y=`TEMPER MAXIMA`), col="red")
```





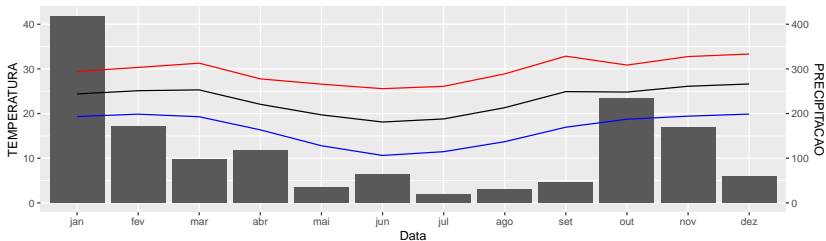
```
dados_meteorologicos %>%  
  select(Data, PRECIPITACAO, starts_with("TEMPER")) %>%  
  pivot_longer(starts_with("TEMPER"), names_to="TEMPER") %>%  
  ggplot(aes(x=Data)) +  
  geom_bar(stat="unique", aes(y=PRECIPITACAO)) +  
  geom_line(aes(y=value, color=TEMPER)) +  
  scale_color_manual(values=c("red", "black", "blue"))
```





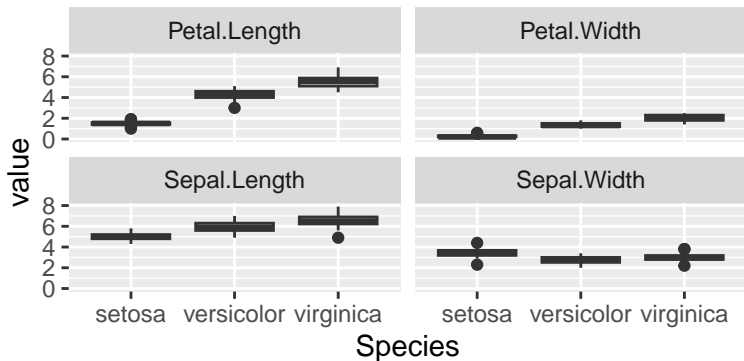
```
dados_meteorologicos %>%
  mutate_at(vars(Data), month, label=T) %>%
  group_by(Data) %>%
  summarise(PRECIPITACAO = sum(PRECIPITACAO),
            `TEMPER MEDIA` = mean(`TEMPER MEDIA`),
            `TEMPER MINIMA` = mean(`TEMPER MINIMA`),
            `TEMPER MAXIMA` = mean(`TEMPER MAXIMA`)) %>%

  ggplot(aes(x=Data)) +
  scale_y_continuous(name="TEMPERATURA",
                    sec.axis = sec_axis(~.*10, name="PRECIPITACAO")) +
  geom_col(aes(y=PRECIPITACAO/10)) +
  geom_line(aes(y=`TEMPER MEDIA`, group=1)) +
  geom_line(aes(y=`TEMPER MINIMA`, group=1, col="blue")) +
  geom_line(aes(y=`TEMPER MAXIMA`, group=1, col="red"))
```



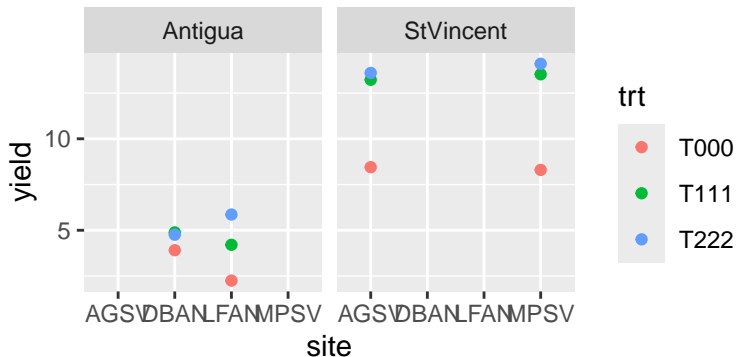


```
iris %>%
  pivot_longer(-Species) %>%
  ggplot(aes(x=Species, y=value)) +
  geom_boxplot() +
  facet_wrap(~name)
```



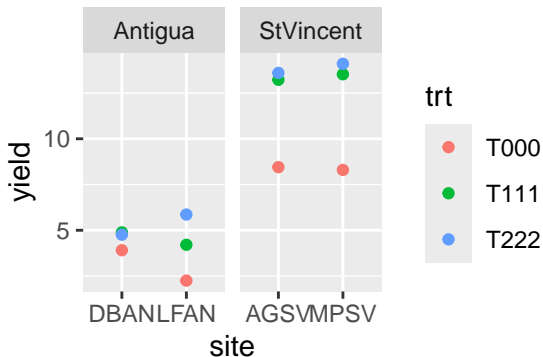


```
agridat::caribbean.maize %>%
  filter(site %in% c("DBAN", "LFAN", "MPSV", "AGSV")) %>%
  filter(trt %in% c("T000", "T111", "T222")) %>%
  group_by(isle, site, trt) %>%
  summarise_at(vars(yield), mean) %>%
  ggplot(aes(x=site, col=trt, y=yield)) +
  geom_point() +
  facet_wrap(~isle)
```



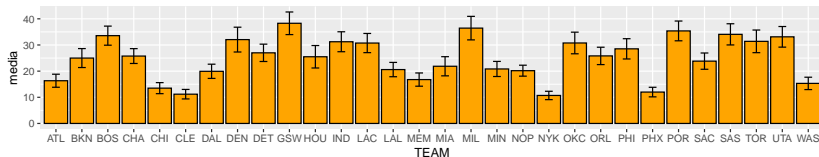


```
agridat::caribbean.maize %>%  
  filter(site %in% c("DBAN", "LFAN", "MPSV", "AGSV")) %>%  
  filter(trt %in% c("T000", "T111", "T222")) %>%  
  group_by(isle, site, trt) %>%  
  summarise_at(vars(yield), mean) %>%  
  ggplot(aes(x=site, col=trt, y=yield)) +  
  geom_point() +  
  facet_wrap(~isle, scales = "free_x")
```



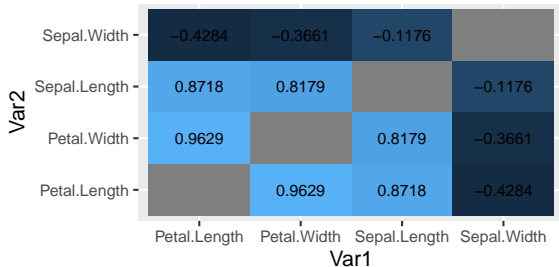


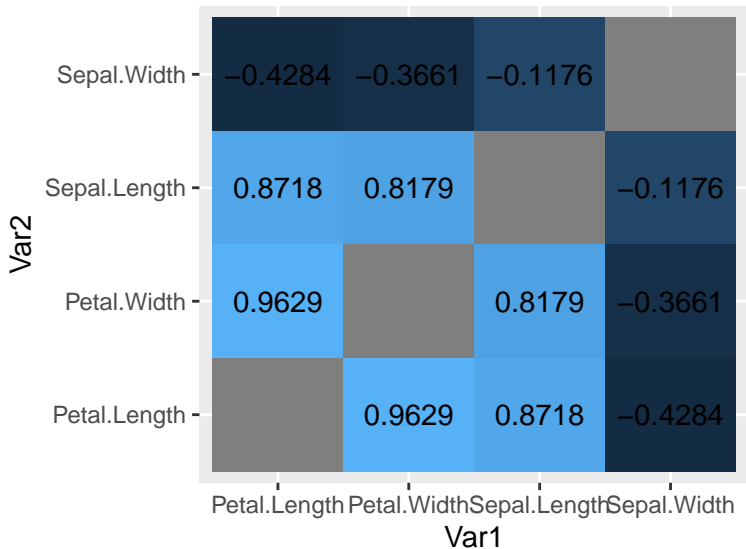
```
library(datasetsICR)
data(NBA.game)
NBA.game %>%
  group_by(TEAM) %>%
  summarise_at(vars(W),
                list(media=mean,
                     erro=\(x) sd(x)/sqrt(length(x)))) %>%
  ggplot(aes(x=TEAM, y=media)) +
  geom_col(fill="orange", col="black") +
  geom_errorbar(aes(ymin=media-erro, ymax=media+erro),
                width=.3, col="black")
```





```
iris %>%
  select(-Species) %>%
  cor() %>%
  as.data.frame() %>%
  rownames_to_column("Var1") %>%
  pivot_longer(-Var1, names_to="Var2", values_to="Cor") %>%
  mutate(Cor = ifelse(Var1==Var2, NA, Cor)) %>%
  ggplot(aes(x=Var1, y=Var2, fill=Cor)) +
  geom_tile(show.legend=F) +
  geom_text(aes(label=round(Cor, 4)), size=2) +
  theme_grey(base_size = 8)
```

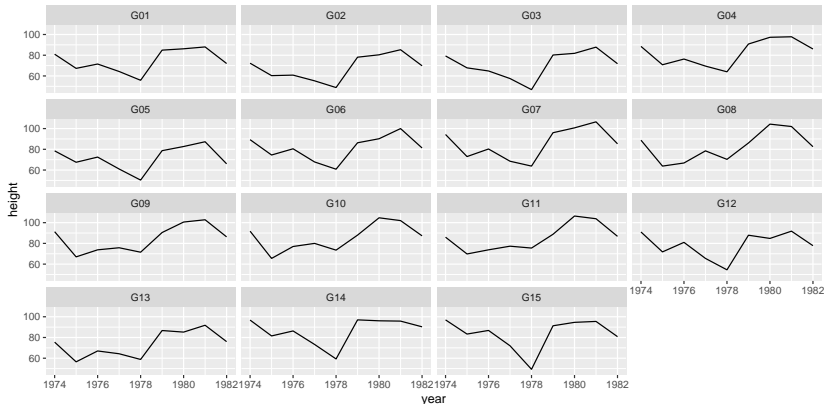




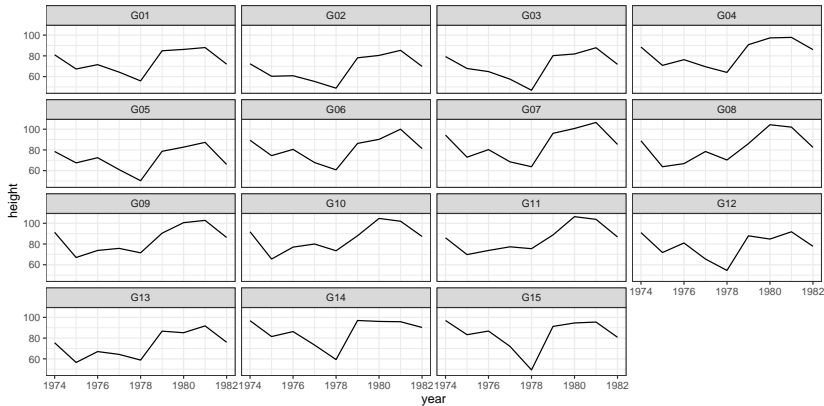




```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line()
```



```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line() +  
  theme_bw()
```

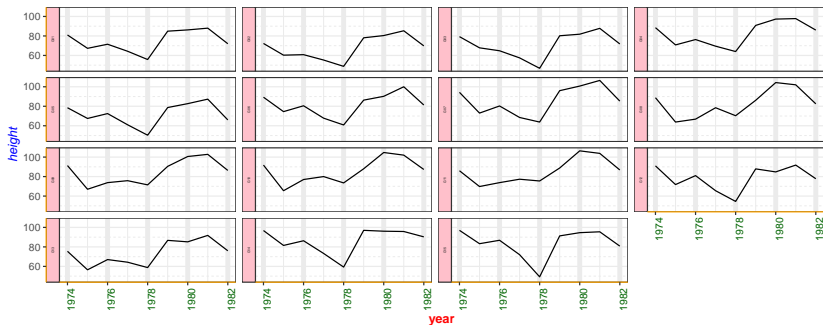




```

agridat::aastveit.barley.height %>%
  ggplot(aes(x=year, y=height)) +
  facet_wrap(~gen, strip.position = "left") +
  geom_line() +
  theme_bw() +
  theme(axis.title.x = element_text(color="red", face="bold"),
        axis.title.y = element_text(color="blue", face="italic"),
        axis.text.x = element_text(color="darkgreen", angle=90),
        strip.background = element_rect(fill="pink"),
        strip.text = element_text(size=3),
        axis.line = element_line(color="orange"),
        panel.grid.major.x = element_line(linewidth=2),
        panel.grid.minor.y = element_line(linetype="dashed"))

```



# Exercícios

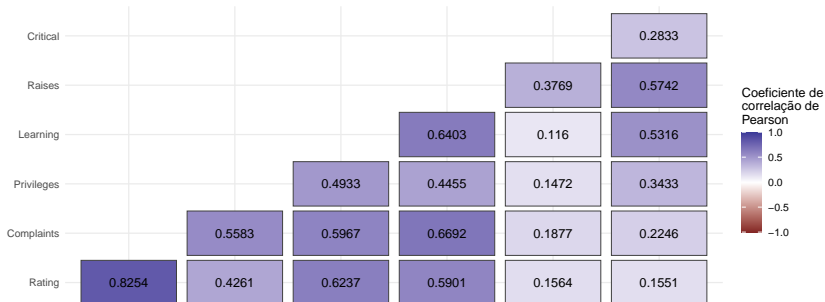
## Exercício



Personalizar o gráfico a seguir:

```
cor(attitude) %>%  
  as.data.frame() %>%  
  rownames_to_column("Var1") %>%  
  pivot_longer(-Var1, names_to="Var2") %>%  
  ggplot(aes(x=Var1, y=Var2, fill=value)) +  
  geom_tile()
```

Exemplo:



# Exemplos

## Conjuntos de dados do governo



<https://dados.gov.br/dados/conjuntos-dados>

# Plantas alternativas a exóticas invasoras



Fonte: <https://dados.gov.br/dados/conjuntos-dados/plantas-alternativas-as-plantas-exoticas-invasoras-rj>

```
library(readxl)
read_xlsx(
  paste0("../dados/listagem-de-plantas-alternativas-as-",
          "plantas-exoticas-invasoras-listadas-para-o-",
          "estado-do-rio-de.xlsx"),
  range="A7:L268")

## New names:
## * `Nome científico` -> `Nome científico...5`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`
## * `Nome científico` -> `Nome científico...7`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`
## * `Nome científico` -> `Nome científico...9`
## * `Nativa (N) ou Exótica (E)` -> `Nativa (N) ou Exótica (E)..`

## # A tibble: 261 x 12
##   `Código de uso` `Família EEI` `Nome científico EEI`
##   <chr>          <chr>          <chr>
```



# Plantas alternativas a exóticas invasoras



Fonte: <https://dados.gov.br/dados/conjuntos-dados/plantas-alternativas-as-plantas-exoticas-invasoras-rj>

```
library(readxl)
invasoras <- read_xlsx(
  paste0("../dados/listagem-de-plantas-alternativas-as-",
    "plantas-exoticas-invasoras-listadas-para-o-",
    "estado-do-rio-de.xlsx"),
  col_names=c("Codigo", "Familia_EEI", "NomeCient_EEI",
    "NomePop_EEI",
    "NomeCient_Ombro", "NouE_Ombro",
    "NomeCient_Semidec", "NouE_Semidec",
    "NomeCient_InfMar", "NouE_InfMar",
    "Ecossist_EEI", "Categoria"),
  range="A8:L268")
```

# Plantas alternativas a exóticas invasoras



```
invasoras %>%  
  pivot_longer(c(ends_with("Ombro"),  
                 ends_with("Semidec"),  
                 ends_with("InfMar"))) %>%  
  separate(name, c("Var", "Ecosystem"))
```

```
## # A tibble: 1,566 x 9
```

```
##   Codigo      Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem  
##   <chr>      <chr>      <chr>      <chr>      <chr>  
## 1 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 2 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 3 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 4 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 5 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 6 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 7 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 8 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 9 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações  
## 10 ornamental Acanthaceae Asystasia ga~ asistásia,~ Formações
```

```
## #> # A tibble: 1,566 x 9
```

## Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar")) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var)
```

```
## # A tibble: 354 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ assistásia,~ Formações P~
## 4 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 5 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 6 orname~ Acanthaceae Brillantaisi~ erva-de-bi~ Floresta Om~
## 7 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 8 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 9 orname~ Acanthaceae Thunbergia a~ amarelinha~ Floresta Om~
## 10 orname~ Acanthaceae Thunbergia g~ tumbérgia-- Floresta Om~
```

# Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar"))) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var) %>%
  unnest_longer(c(NomeCient, NouE))
```

```
## # A tibble: 783 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <d
## 1 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 4 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 5 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 6 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 7 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 8 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 9 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
```

## Plantas alternativas a exóticas invasoras



```
invasoras %>%
  pivot_longer(c(ends_with("Ombro"),
                 ends_with("Semidec"),
                 ends_with("InfMar"))) %>%
  separate(name, c("Var", "Ecosystem")) %>%
  pivot_wider(id_cols=c(Codigo,ends_with("EEI"),Categoria,Ecosystem),
              names_from=Var) %>%
  unnest_longer(c(NomeCient, NouE)) %>%
  filter(!is.na(NomeCient), !is.na(NouE))
```

```
## # A tibble: 504 x 9
```

```
##   Codigo  Familia_EEI NomeCient_EEI NomePop_EEI Ecosystem_EEI Categoria
##   <chr>   <chr>         <chr>         <chr>         <chr>         <chr>
## 1 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 2 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 3 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 4 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 5 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 6 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 7 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
## 8 orname~ Acanthaceae Asystasia ga~ asistásia,~ Formações P~
```



```
group_by(NomeCient_EEI) %>%
```

```
ungroup() %>%
```

geom\_col() +

```
theme(axis.text.x = element_text(angle=90, hjust=1, vjust=.5))
```





```
theme(axis.text.x = element_text(angle=90, hjust=1, vjust=.5))
```



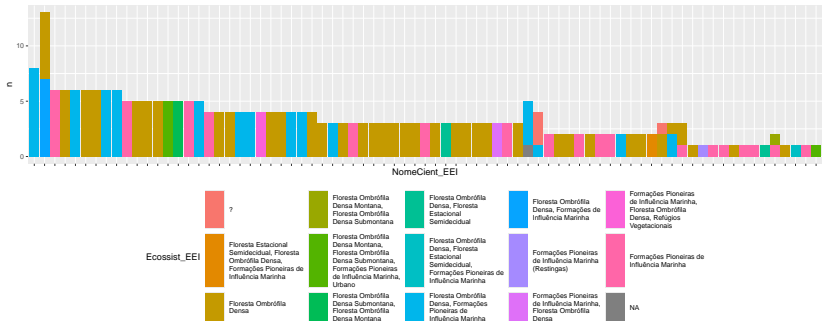


# Plantas alternativas a exóticas invasoras

```

invasoras %>%
  group_by(NomeCient_EEI, Ecosystem_EEI) %>%
  count() %>%
  ungroup() %>%
  mutate(NomeCient_EEI = fct_reorder(NomeCient_EEI, desc(n))) %>%
  mutate_at(vars(Ecosystem_EEI), str_wrap, width=22) %>%
  ggplot(aes(x=NomeCient_EEI, y=n, fill=Ecosystem_EEI)) +
  guides(fill=guide_legend(nrow=3)) +
  geom_col() +
  theme_grey(base_size = 8) +
  theme(axis.text.x = element_blank(),
        legend.position = "bottom")

```





## Plantas alternativas a exóticas invasoras



```
invasoras %>%  
  group_by(NomeCient_EEI, Ecossist_EEI) %>%  
  count() %>%  
  group_by(NomeCient_EEI) %>%  
  count() %>%  
  arrange(desc(n))
```

```
## # A tibble: 77 x 2  
## # Groups:   NomeCient_EEI [77]  
##   NomeCient_EEI      n  
##   <chr>          <int>  
## 1 Acacia angustissima (Mill.) Kuntze      2  
## 2 Acacia mangium Wild      2  
## 3 Dieffenbachia amoena hort. ex Gentil    2  
## 4 Elaeis guineensis Jacq.      2  
## 5 Melinis minutiflora P. Beauv.      2  
## 6 Mimusops commersonii (G. Don) Engl.    2  
## 7 Thunbergia grandiflora Roxb.      2  
## 8 Acacia auriculiformis A.Cunn. ex Benth.  1  
## 9 Aroca americana L.      1
```



## Plantas alternativas a exóticas invasoras

```
ecossist <- invasoras %>%  
  mutate_at(vars(Ecossist_EEI), replace_na, "?") %>%  
  group_by(NomeCient_EEI, Ecossist_EEI) %>%  
  count() %>%  
  group_by(NomeCient_EEI) %>%  
  arrange(Ecossist_EEI) %>%  
  summarise(Ecossist_EEI = (function(x) {  
    if(length(x)==1) return(x)  
    else {  
      if(is.na(x[1])) return(x[2])  
      if(is.na(x[2])) return(x[1])  
      if(grepl(x[1], x[2])) return(x[2])  
      if(grepl(x[2], x[1])) return(x[1])  
      return(paste(x, collapse=", "))  
    }  
  }  
  )(Ecossist_EEI),  
  n = sum(n)) %>%  
  mutate_at(vars(Ecossist_EEI), ~ifelse(  
    .=="Formações Pioneiras de Influência Marinha, Floresta Ombrófila D  
    "Floresta Ombrófila Densa, Formações Pioneiras de Influência Marinh
```

# Plantas alternativas a exóticas invasoras



```
ecossist %>%
  mutate(NomeCient_EEI = fct_reorder(NomeCient_EEI, desc(n))) %>%
  mutate_at(vars(Ecossist_EEI), str_wrap, width=22) %>%
  ggplot(aes(x=NomeCient_EEI, y=n, fill=Ecossist_EEI)) +
  guides(fill=guide_legend(nrow=3)) +
  geom_col() +
  theme_grey(base_size = 8) +
  theme(axis.text.x = element_blank(),
        legend.position = "bottom")
```





# Plantas alternativas a exóticas invasoras

```
ecossist %>%
  mutate(NomeCient_EEI = fct_reorder(NomeCient_EEI, desc(n))) %>%
  mutate_at(vars(Ecossist_EEI), str_wrap, width=22) %>%
  mutate(Ecossist_EEI = fct_reorder(Ecossist_EEI, desc(n))) %>%
  ggplot(aes(x=NomeCient_EEI, y=n, fill=Ecossist_EEI)) +
  guides(fill=guide_legend(nrow=3)) +
  geom_col() +
  theme_grey(base_size = 8) +
  theme(axis.text.x = element_blank(),
        legend.position = "bottom")
```





```
library(readr)
```

```
fomentoRural <- read_csv("https://aplicacoes.mds.gov.br/sagi/ser
```

```
head(fomentoRural)
```

```
## # A tibble: 6 x 1,077
```

```
##   paa_qtd_agricul_familiar_modal~1 qtd_pes_pbf_idade_7_~2 bol
```

```
##   <lgl>                                <dbl> <lgl>
```

```
## 1 NA                                101 NA
```

```
## 2 NA                                105 NA
```

```
## 3 NA                                104 NA
```

```
## 4 NA                                94 NA
```

```
## 5 NA                                128 NA
```

```
## 6 NA                                90 NA
```

```
## # i abbreviated names: 1: paa_qtd_agricul_familiar_modal_form
```

```
## #   2: qtd_pes_pbf_idade_7_a_15_sexo_feminino_i,
```

```
## #   3: bolsa_atleta_qtd_atletas_ens_fund_incomp_16_18_anos_i
```

```
## # i 1,074 more variables: bolsa_atleta_qtd_atletas_sexo_61m_c
```

```
## #   bolsa_atleta_qtd_atletas_ens_superior_compl_14_15_anos_i
```

## Programa de Fomento Rural



```
fomentoRural %>%  
  select_if(~!all(is.na(.))) %>%  
  select_if(~!all(==0))
```

```
## # A tibble: 559 x 265  
##   qtd_pes_pbf_idade_7_a_15_sexo~1 igd_pbf_qtd_total_pu~2 qtd  
##                                     <dbl>               <dbl>  
## 1                                101                2340  
## 2                                105                2340  
## 3                                104                3420  
## 4                                 94                2655  
## 5                                128               11724  
## 6                                 90                8478  
## 7                                 78                3169  
## 8                                350               16477  
## 9                                354               16543  
## 10                               345               16543  
## # i 549 more rows  
## # i abbreviated names: 1: qtd_pes_pbf_idade_7_a_15_sexo_62min  
## # 2: igd_pbf_qtd_total_publica_saudefam_62min
```



```
fomentoRural %>%  
  select_if(~mean(is.na(.)| .==0)<.5) %>%  
  select(mes_ano, matches(paste0(  
    "qtd\\_pes\\_(cad\\_nao\\_)?pbfb",  
    "\\_idade\\_\\.+\\_sexo.+\\_i"))) )
```

```
## # A tibble: 559 x 53
```

```
##   mes_ano          qtd_pes_pbf_idade_7_a_15_sexo_fe~1 qtd  
##   <dtm>                                <dbl>  
## 1 2024-03-01 00:00:00          101  
## 2 2024-02-01 00:00:00          105  
## 3 2024-05-01 00:00:00          104  
## 4 2024-05-01 00:00:00           94  
## 5 2024-03-01 00:00:00          128  
## 6 2024-01-01 00:00:00           90  
## 7 2024-01-01 00:00:00           78  
## 8 2024-01-01 00:00:00          350  
## 9 2024-02-01 00:00:00          354  
## 10 2024-03-01 00:00:00          3463  
## # i 549 more rows
```



```
fomentoRural %>%
  select_if(~mean(is.na(.)|.==0)<.5) %>%
  select(mes_ano, matches(
    "qtd\\_pes\\_(cad\\_nao\\_)?pbf\\_idade\\_\\.+\\_sexo.+\\_i")) %>%
  group_by(mes_ano) %>%
  summarise_all(sum, na.rm=T)
```

```
## # A tibble: 5 x 53
##   mes_ano          qtd_pes_pbf_idade_7_a_15_sexo_fem~1 qtd_pes_ca
##   <dtm>                                <dbl>
## 1 2024-01-01 00:00:00                                33980
## 2 2024-02-01 00:00:00                                48934
## 3 2024-03-01 00:00:00                                41020
## 4 2024-04-01 00:00:00                                65027
## 5 2024-05-01 00:00:00                                86640
## # i abbreviated names: 1: qtd_pes_pbf_idade_7_a_15_sexo_feminino_i,
## #   2: qtd_pes_cad_nao_pbf_idade_16_a_17_sexo_masculino_i
## # i 50 more variables: qtd_pes_cad_nao_pbf_idade_60_a_64_sexo_femini
## #   qtd_pes_pbf_idade_50_a_54_sexo_feminino_i <dbl>,
## #   qtd_pes_cad_nao_pbf_idade_60_a_64_sexo_masculino_i <dbl>,
## #   qtd_pes_pbf_idade_16_a_17_sexo_masculino_i <dbl>,
```





```
fr <- fomentoRural %>%
  select_if(~mean(is.na(.)|.==0)<.5) %>%
  select(mes_ano, matches(
    "qtd\\_pes\\_(cad\\_nao\\_)?pbf\\_idade\\_\\.+\\_sexo.+\\_i")) %>%
  group_by(mes_ano) %>%
  summarise_all(sum, na.rm=T) %>%
  pivot_longer(-mes_ano)
```

```
head(fr)
```

```
## # A tibble: 6 x 3
```

```
##   mes_ano          name
##   <dtm>          <chr>
```

```
## 1 2024-01-01 00:00:00 qtd_pes_pbf_idade_7_a_15_sexo_feminino_i
```

```
## 2 2024-01-01 00:00:00 qtd_pes_cad_nao_pbf_idade_16_a_17_sexo_masculi
```

```
## 3 2024-01-01 00:00:00 qtd_pes_cad_nao_pbf_idade_60_a_64_sexo_feminin
```

```
## 4 2024-01-01 00:00:00 qtd_pes_pbf_idade_50_a_54_sexo_feminino_i
```

```
## 5 2024-01-01 00:00:00 qtd_pes_cad_nao_pbf_idade_60_a_64_sexo_masculi
```

```
## 6 2024-01-01 00:00:00 qtd_pes_pbf_idade_16_a_17_sexo_masculino_i
```



```
fr <- fr %>%  
  mutate(pbf = str_extract(name, "(nao\\_)?pbf"),  
         idade=str_extract(  
           name,  
           "idade\\_(maior\\_que\\_)?\\d+(\\_\\[ea]\\_\\_\\d+)?"),  
         sexo =str_extract(name, "sexo\\_\\[^\\_\\_]+ino"))  
  
head(fr)
```

```
## # A tibble: 6 x 6
```

##	mes_ano		name	value	pbf
##	<dtm>		<chr>	<dbl>	<chr>
## 1	2024-01-01 00:00:00	qtd_pes_pbf_idade_7_a_15_sexo_fem~		33980	pbf
## 2	2024-01-01 00:00:00	qtd_pes_cad_nao_pbf_idade_16_a_17~		9534	nao_~
## 3	2024-01-01 00:00:00	qtd_pes_cad_nao_pbf_idade_60_a_64~		8224	nao_~
## 4	2024-01-01 00:00:00	qtd_pes_pbf_idade_50_a_54_sexo_fe~		24579	pbf
## 5	2024-01-01 00:00:00	qtd_pes_cad_nao_pbf_idade_60_a_64~		21729	nao_~
## 6	2024-01-01 00:00:00	qtd_pes_pbf_idade_16_a_17_sexo_ma~		17702	pbf



```
fr <- fr %>%  
  mutate(pbf = factor(pbf, levels=c("pbf", "nao_pbf"),  
                        labels=c("PBF", "Não-PBF")),  
         idade=str_replace_all(str_remove(idade, "idade\\_"),  
                               "\\_", " "),  
         sexo =str_to_title(str_remove(sexo, "sexo\\_"))) %>%  
  mutate_at(vars(idade, sexo), factor)  
  
levels(fr$idade)
```

```
## [1] "0 e 4"          "16 a 17"        "18 a 24"        "25 a 34"  
## [6] "40 a 44"        "45 a 49"        "5 a 6"          "50 a 54"  
## [11] "60 a 64"        "7 a 15"         "maior que 65"
```



```
fr <- fr %>%  
  mutate(idade1 = parse_number(str_extract(idade, "\\d+"))) %>%  
  mutate_at(vars(idade, sexo), factor) %>%  
  mutate(idade = fct_reorder(idade, idade1))  
  
levels(fr$idade)
```

```
## [1] "0 e 4"          "5 a 6"          "7 a 15"         "16 a 17"  
## [6] "25 a 34"        "35 a 39"        "40 a 44"        "45 a 49"  
## [11] "55 a 59"        "60 a 64"        "maior que 65"
```

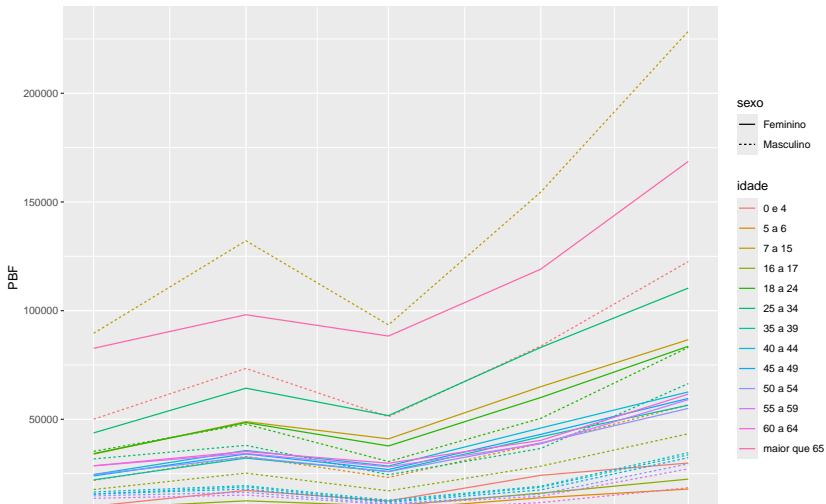


```
fr <- fr %>%  
  select(mes_ano, pbf, idade, sexo, name, value) %>%  
  arrange(mes_ano, pbf, idade, sexo) %>%  
  pivot_wider(id_cols=c(mes_ano, idade, sexo), names_from=pbf)  
  
head(fr)
```

```
## # A tibble: 6 x 5  
##   mes_ano          idade  sexo      PBF `Não-PBF`  
##   <dtm>          <fct>  <fct>    <dbl>    <dbl>  
## 1 2024-01-01 00:00:00 0 e 4  Feminino  10059    47620  
## 2 2024-01-01 00:00:00 0 e 4  Masculino 50038    10887  
## 3 2024-01-01 00:00:00 5 a 6  Feminino   6789    21025  
## 4 2024-01-01 00:00:00 5 a 6  Masculino 21885     7325  
## 5 2024-01-01 00:00:00 7 a 15 Feminino  33980    85698  
## 6 2024-01-01 00:00:00 7 a 15 Masculino 89566    36549
```



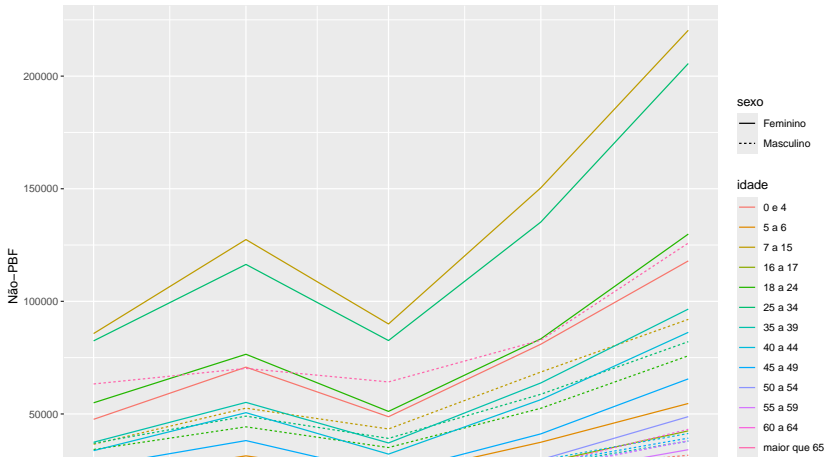
```
fr %>%  
  ggplot(aes(x=mes_ano, col=idade, linetype=sexo, y=PBF)) +  
  geom_line()
```





```
fr %>%
```

```
  ggplot(aes(x=mes_ano, col=idade, linetype=sexo, y=`Não-PBF`))  
  geom_line()
```



# Exercícios



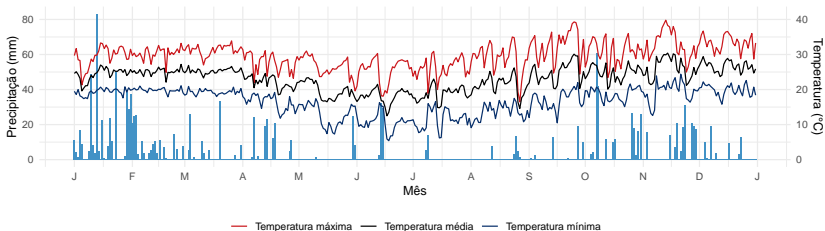
# Exercício 1



Personalizar o gráfico a seguir:

```
dados_meteorologicos %>%  
  select(Data, PRECIPITACAO, starts_with("TEMPER")) %>%  
  pivot_longer(starts_with("TEMPER")) %>%  
  ggplot(aes(x=Data)) +  
    geom_bar(stat="unique", aes(y=PRECIPITACAO)) +  
    geom_line(aes(y=value, color=name))
```

Exemplo:



## Exercício 2

