

Manipulação e Apresentação de Dados

Guilherme Bovi Ambrosano



▶ Material do curso

▶ broom

▶ purrr

▶ ggplot2

▶ Exemplos

▶ Exercícios

Material do curso



https://github.com/guilherme-ambrosano/Curso_R_2024

The screenshot shows the GitHub interface for the repository 'Curso_R_2024'. The repository is public and has 1 branch (main) and 0 tags. The commit history shows 11 commits. The most recent commit is by guilherme-ambrosano, titled 'Adicionando o exemplo das bolsas', made 2 days ago. The commit history table is as follows:

Commit	Description	Time ago
Dia1	Colocando os exercicios antes da familia apply	2 days ago
Dia2	Primeiro commit	5 days ago
Dia3	Adicionando o exemplo das bolsas	2 days ago
.gitignore	Primeiro commit	5 days ago

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
```

```
library(readxl)
excel_sheets(paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx"))
```

```
## [1] "FEV-2023" "MAR-2023" "ABR-2023" "MAI-2023" "JUN-2023" "
## [7] "AGO-2023" "SET-2023" "OUT-2023" "NOV-2023" "DEZ-2023"
```



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

map(excel_sheets(arquivo_xlsx),
  \(planilha) read_xlsx(arquivo_xlsx, planilha))
```

```
## [[1]]
## # A tibble: 53 x 2
##   BOLSISTA                                `VALOR DA BOLSA`
##   <chr>                                <chr>
## 1 MIKAEL GABRIEL TOMAZ DA SILVA        R$ 500,00
## 2 ASHLEY LARISSA GOMES DE MEDEIROS     R$ 500,00
## 3 ADENILSON DA PENHA RIBEIRO           R$ 500,00
## 4 BRENDÓ DIAS DO NASCIMENTO            R$ 500,00
## 5 GUSTAVO AURÉLIO LINHARES DE MAGALHÃES R$ 500,00
## 6 GRAZIELLY MARQUES NASCIMENTO SILVA   R$ 500,00
## 7 MYLENA HUK GAZEL                     R$ 500,00
## 8 LUY SOUZA DE SOUZA                   R$ 500,00
## 9 VANESSA CORDEIRO SENA                 R$ 500,00
```



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

map_dfr(excel_sheets(arquivo_xlsx),
  \(planilha) read_xlsx(arquivo_xlsx, planilha))
```

```
## Error in `dplyr::bind_rows()`:
```

```
## ! Can't combine `..1$VALOR DA BOLSA` <character> and `..8$VAL
```



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

map(excel_sheets(arquivo_xlsx),
  \(planilha) {
    unique(read_xlsx(arquivo_xlsx, planilha)$`VALOR DA BOLSA`)
  })
```

```
## [[1]]
## [1] "R$ 500,00"
##
## [[2]]
## [1] "R$ 500,00"
##
## [[3]]
## [1] "R$ 500,00"
##
## [[4]]
## [1] "R$ 500,00"
##
## [[5]]
## [1] "R$ 500,00" NA
##
## [[6]]
## [1] "R$ 500,00"
##
## [[7]]
## [1] "R$ 500,00"
##
## [[8]]
## [1] 500
##
## [[9]]
## [1] 500
##
## [[10]]
## [1] 500
##
```



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

map_dfr(excel_sheets(arquivo_xlsx),
  \(planilha) {
    bolsas_mes <- read_xlsx(arquivo_xlsx, planilha)
    if(class(bolsas_mes$`VALOR DA BOLSA`)=="numeric") {
      bolsas_mes <- bolsas_mes %>%
        mutate_at(vars(`VALOR DA BOLSA`), ~paste("R$", .))
    }
    return(bolsas_mes)
  })
```

```
## # A tibble: 707 x 2
```

##	BOLSISTA	`VALOR DA BOLSA`
##	<chr>	<chr>
##	1 MIKAEL GABRIEL TOMAZ DA SILVA	R\$ 500,00
##	2 ASHLEY LARISSA GOMES DE MEDEIROS	R\$ 500,00
##	3 ADENILSON DA PENHA RIBEIRO	R\$ 500,00
##	4 BRENDON DIAS DO NASCIMENTO	R\$ 500,00
##	5 GUSTAVO ANDRÉ LOPES DE MACHADO	R\$ 500,00



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

map_dfr(excel_sheets(arquivo_xlsx),
  \(planilha) {
    bolsas_mes <- read_xlsx(arquivo_xlsx, planilha)
    if(class(bolsas_mes$`VALOR DA BOLSA`)=="numeric") {
      bolsas_mes <- bolsas_mes %>%
        mutate_at(vars(`VALOR DA BOLSA`), ~paste("R$", .))
    }
    return(bolsas_mes)
  }) %>%
  mutate_at(vars(`VALOR DA BOLSA`), parse_number,
    locale=locale(decimal=","))
```

```
## # A tibble: 707 x 2
##   BOLSISTA                                `VALOR DA BOLSA`
##   <chr>                                <dbl>
## 1 MIKAEL GABRIEL TOMAZ DA SILVA          500
## 2 ASHLEY LARISSA GOMES DE MEDEIROS        500
## 3 ADRIANILSON DA SILVA PEREIRA          500
```



Algun bug no lubridate faz ele retornar 08 pro mês ABR

```
parse_date_time2("ABR-2023", orders="%B-%Y")
```

```
## [1] "2023-08-01 UTC"
```

```
parse_date_time2("AUG-2023", orders="%B-%Y")
```

```
## [1] "2023-08-01 UTC"
```

```
parse_date_time2("APR-2023", orders="%B-%Y")
```

```
## [1] "2023-04-01 UTC"
```



```
arquivo_xlsx <- paste0(
  "../dados/Relação nominal dos alunos - ",
  "Bolsa Extensão - 2023.xlsx")

bolsas <- map_dfr(excel_sheets(arquivo_xlsx),
  \(planilha) {
    read_xlsx(arquivo_xlsx, planilha) %>%
      mutate_at(vars(`VALOR DA BOLSA`),
        as.character) %>%
      mutate(Mes=planilha)
  }) %>%
  mutate_at(vars(`VALOR DA BOLSA`), parse_number,
    locale=locale(decimal=",")) %>%
  mutate_at(vars(Mes), ~ifelse(=="ABR-2023", "APR-2023", .)) %>%
  mutate_at(vars(Mes), parse_date_time2, "%B-%Y")
```



```
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"  "-"
```



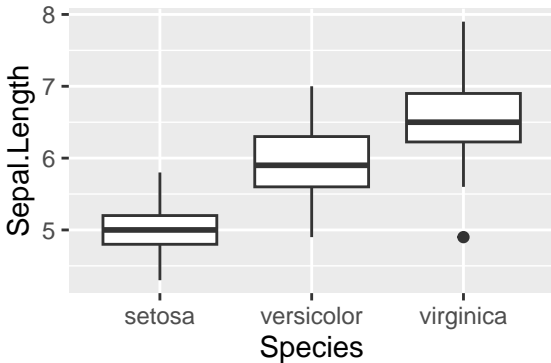
```
map2(c(1,2,3,4), c("a","b","c","d"),  
      \(num, chr) paste0("Número: ", num, "; Caractere: ", chr))
```

```
## [[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

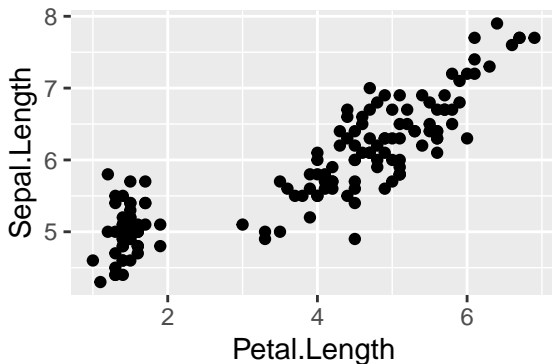


```
library(ggplot2)
iris %>%
  ggplot(aes(x=Species, y=Sepal.Length)) +
  geom_boxplot()
```



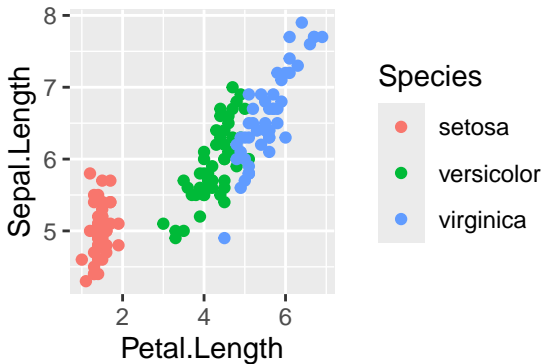


```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length)) +  
  geom_point()
```



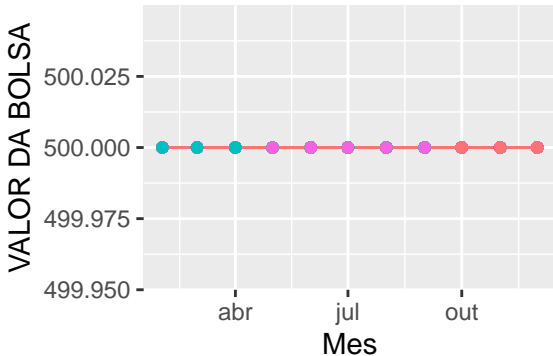


```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length,  
             color=Species)) +  
  geom_point()
```

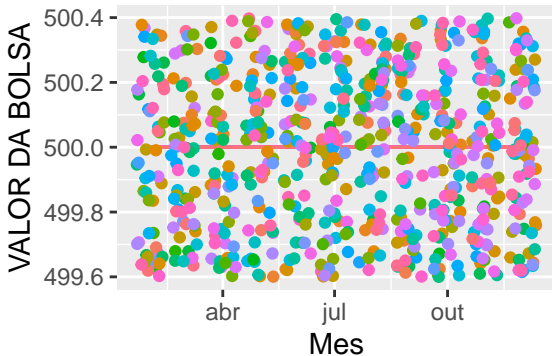




```
bolsas %>%  
  ggplot(aes(x=Mes, y=`VALOR DA BOLSA`, col=BOLSISTA)) +  
  geom_line(show.legend=F) +  
  geom_point(show.legend=F)
```

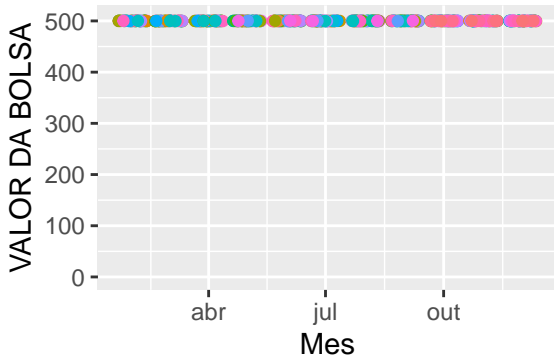


```
bolsas %>%  
  ggplot(aes(x=Mes, y=`VALOR DA BOLSA`, col=BOLSISTA)) +  
  geom_line(show.legend=F) +  
  geom_point(show.legend=F, position = "jitter")
```



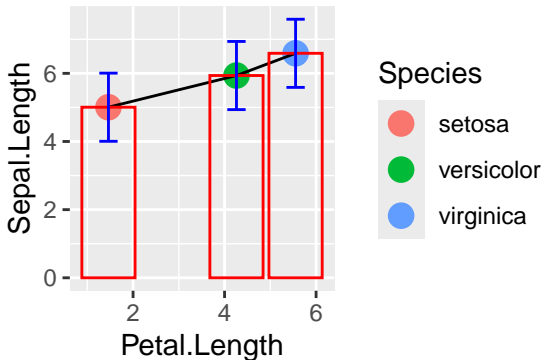


```
bolsas %>%  
  ggplot(aes(x=Mes, y=`VALOR DA BOLSA`, col=BOLSISTA)) +  
  scale_y_continuous(limits=c(0, 505)) +  
  geom_line(show.legend=F) +  
  geom_point(show.legend=F, position = "jitter")
```



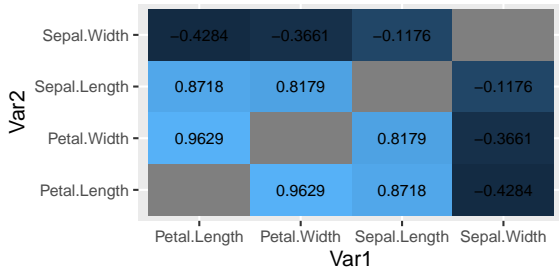


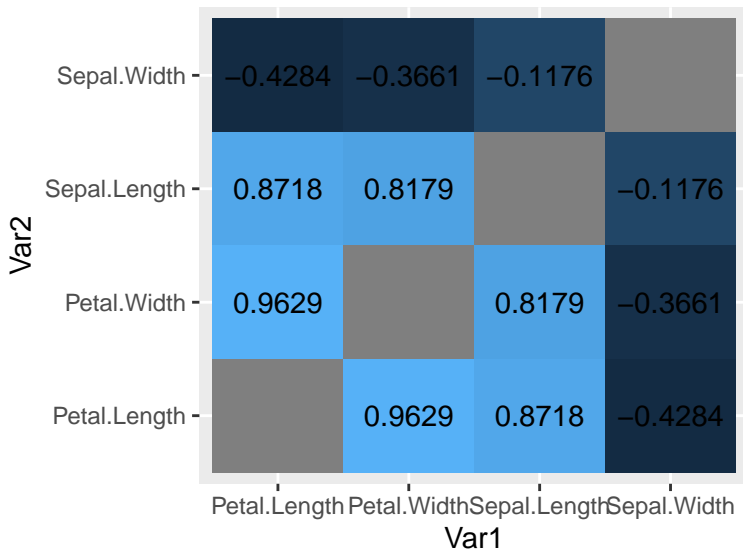
```
iris %>%  
  group_by(Species) %>%  
  summarise_at(vars(Petal.Length, Sepal.Length), mean) %>%  
  ggplot(aes(x=Petal.Length, y=Sepal.Length, col=Species)) +  
  geom_point(size=4) +  
  geom_line(col="black") +  
  geom_col(col="red", fill="transparent") +  
  geom_errorbar(aes(ymax=Sepal.Length+1,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





```
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)
```

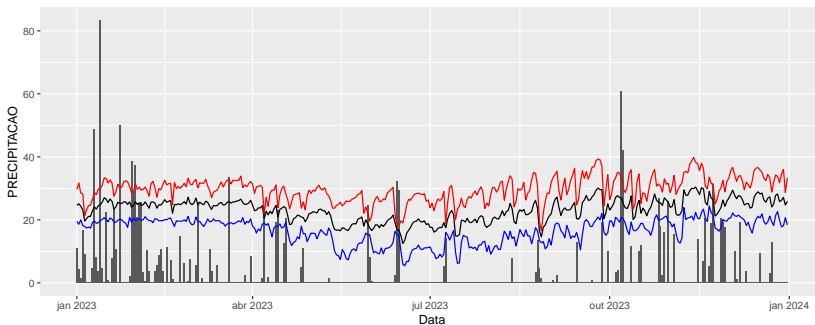




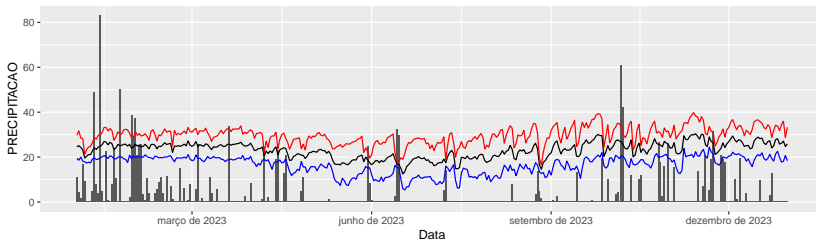


```
dados_meteorologicos <- read_lines(  
  "http://www.leb.esalq.usp.br/leb/exceldados/DCE2023.TXT") %>%  
  as_tibble() %>%  
  mutate(n = nchar(value)) %>%  
  filter(n > 98) %>%  
  select(-n) %>%  
  separate(value,  
    c("No", "ANO", "DIA", "MES", "R.GLOBA",  
      "INSOLACAO", "PRECIPITACAO", "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")
```

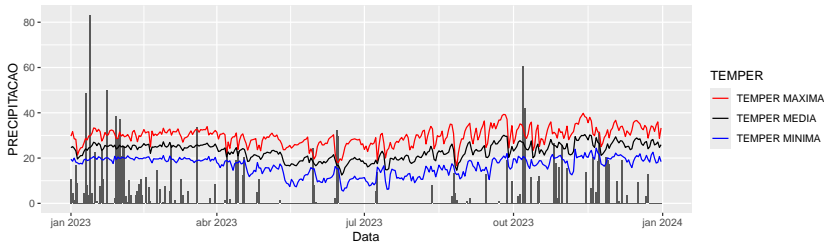


```
dados_meteorologicos %>%
  ggplot(aes(x=Data)) +
  scale_x_date(date_breaks="3 months",
               date_labels="%B de %Y") +
  geom_col(aes(y=PRECIPITACAO)) +
  geom_line(aes(y=`TEMPER MEDIA`)) +
  geom_line(aes(y=`TEMPER MINIMA`, col="blue")) +
  geom_line(aes(y=`TEMPER MAXIMA`, col="red"))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



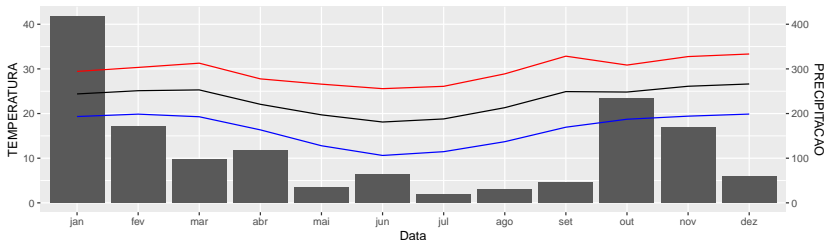


```
dados_meteorologicos %>%  
  select(Data, PRECIPITACAO, starts_with("TEMPER")) %>%  
  pivot_longer(starts_with("TEMPER"), names_to="TEMPER") %>%  
  ggplot(aes(x=Data)) +  
  geom_bar(stat="unique", aes(y=PRECIPITACAO)) +  
  geom_line(aes(y=value, color=TEMPER)) +  
  scale_color_manual(values=c("red",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



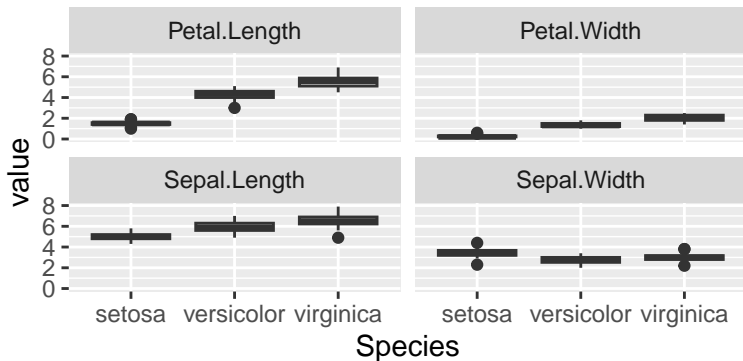


```
dados_meteorologicos %>%
  mutate_at(vars(Data), month, label=T) %>%
  group_by(Data) %>%
  summarise(PRECIPITACAO = sum(PRECIPITACAO),
            `TEMPER MEDIA` = mean(`TEMPER MEDIA`),
            `TEMPER MINIMA` = mean(`TEMPER MINIMA`),
            `TEMPER MAXIMA` = mean(`TEMPER MAXIMA`)) %>%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



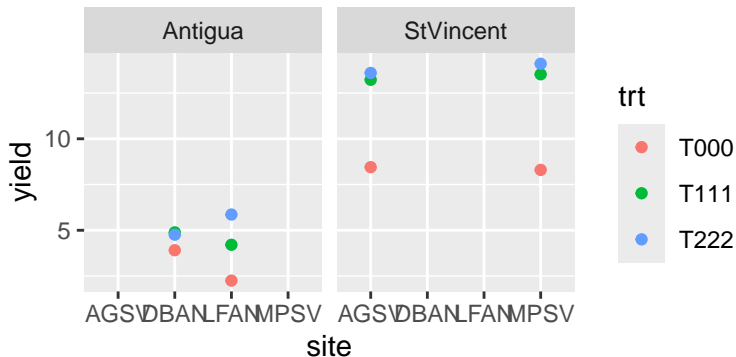


```
iris %>%
  pivot_longer(-Species) %>%
  ggplot(aes(x=Species, y=value)) +
  geom_boxplot() +
  facet_wrap(~name)
```



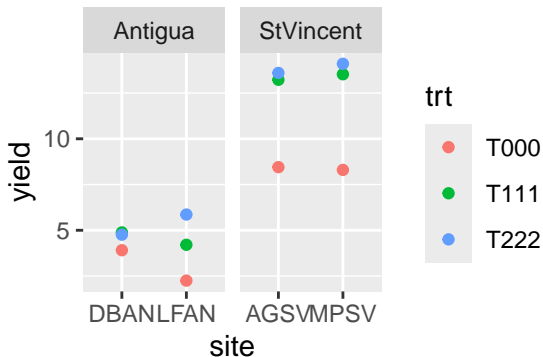


```
agridat::caribbean.maize %>%  
  filter(site %in% c("DBAN", "LFAN", "MPSV", "AGSV")) %>%  
  filter(trt %in% c("T000", "T111", "T222")) %>%  
  group_by(isle, site, trt) %>%  
  summarise_at(vars(yield), mean) %>%  
  ggplot(aes(x=site, col=trt, y=yield)) +  
  geom_point() +  
  facet_wrap(~isle)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



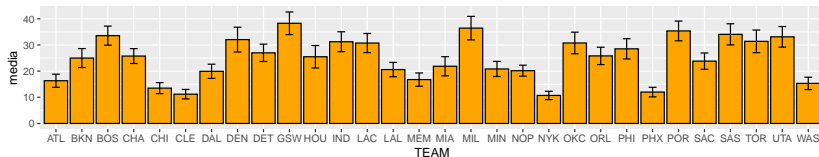


```
agridat::caribbean.maize %>%  
  filter(site %in% c("DBAN", "LFAN", "MPSV", "AGSV")) %>%  
  filter(trt %in% c("T000", "T111", "T222")) %>%  
  group_by(isle, site, trt) %>%  
  summarise_at(vars(yield), mean) %>%  
  ggplot(aes(x=site, col=trt, y=yield)) +  
  geom_point() +  
  facet_wrap(~isle,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



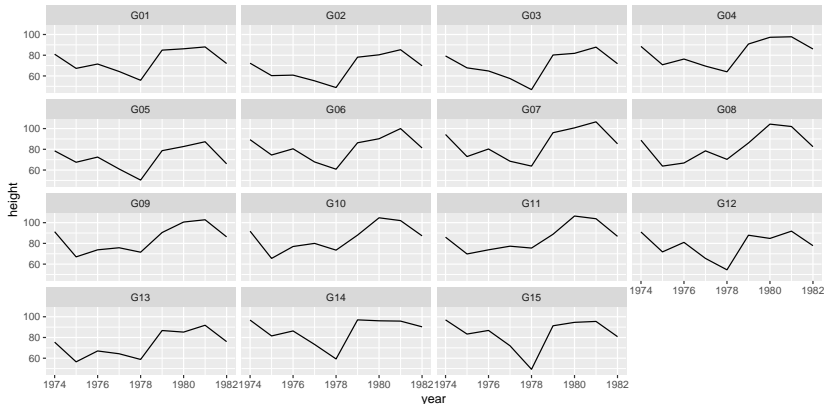


```
library(datasetsICR)
data(NBA.game)
NBA.game %>%
  group_by(TEAM) %>%
  summarise_at(vars(W),
                list(media=mean,
                     erro=\(x) sd(x)/sqrt(length(x)))) %>%
  ggplot(aes(x=TEAM, y=media)) +
  geom_col(fill="orange", col="black") +
  geom_errorbar(aes(ymin=media-erro,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

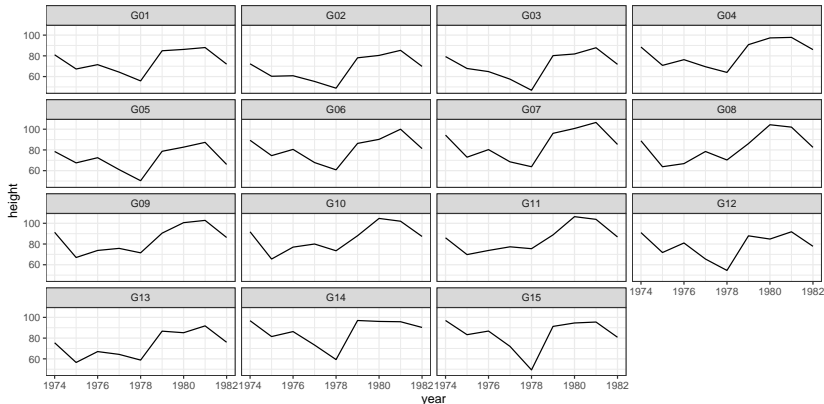




```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line()
```

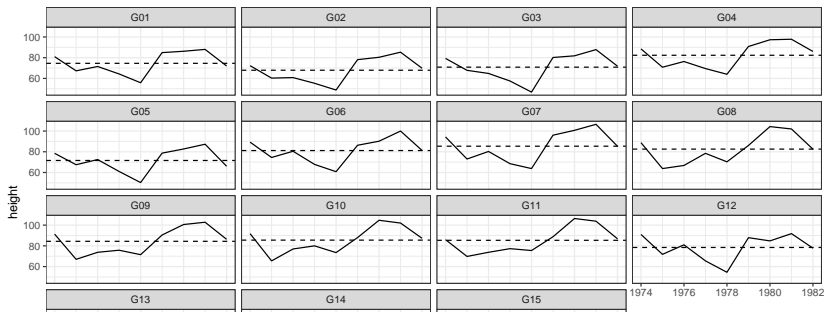


```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line() +  
  theme_bw()
```



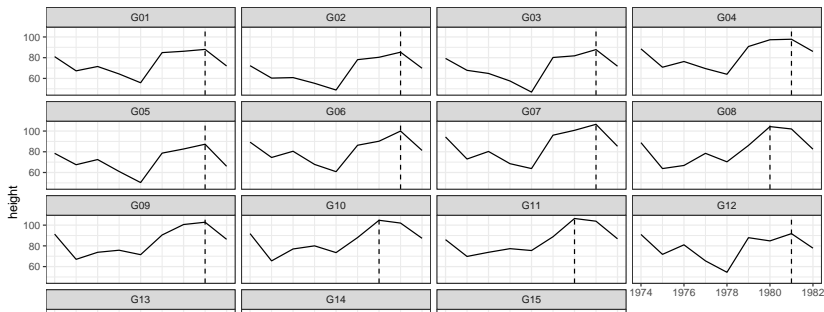


```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line() +  
  geom_hline(  
    data=aggregate(aastveit.barley.height$height,  
                    list(gen=aastveit.barley.height$gen),  
                    mean),  
    aes(yintercept=x), linetype="dashed") +  
  theme_bw()
```





```
agridat::aastveit.barley.height %>%  
  ggplot(aes(x=year, y=height)) +  
  facet_wrap(~gen) +  
  geom_line() +  
  geom_vline(  
    data=aastveit.barley.height %>%  
      group_by(gen) %>%  
      filter(height==max(height)),  
    aes(xintercept=year), linetype="dashed") +  
  theme_bw()
```

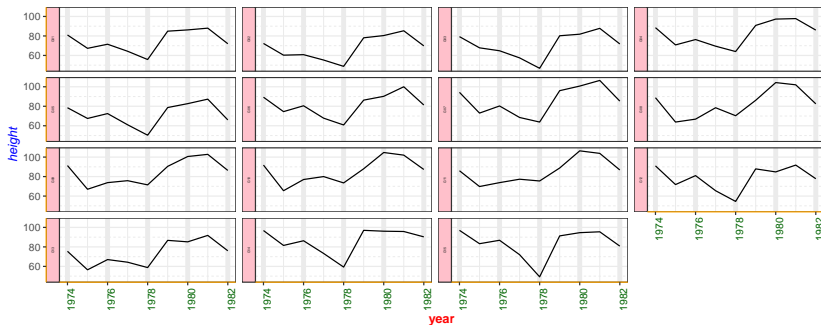




```

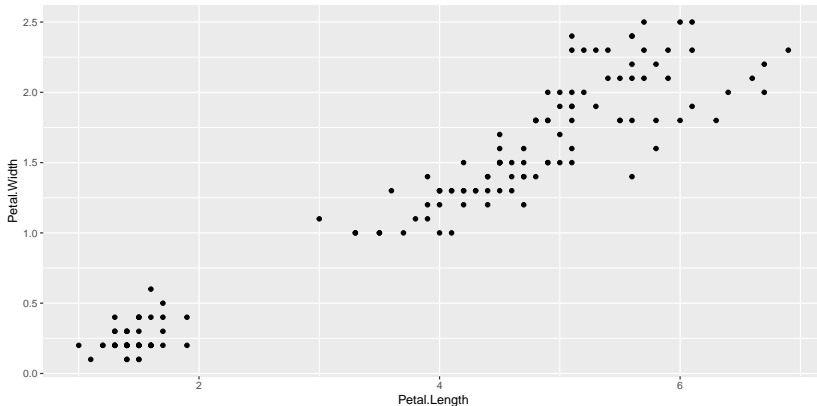
agridat::aastveit.barley.height %>%
  ggplot(aes(x=year, y=height)) +
  facet_wrap(~gen, strip.position = "left") +
  geom_line() +
  theme_bw() +
  theme(axis.title.x = element_text(color="red", face="bold"),
        axis.title.y = element_text(color="blue", face="italic"),
        axis.text.x = element_text(color="darkgreen", 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"))

```





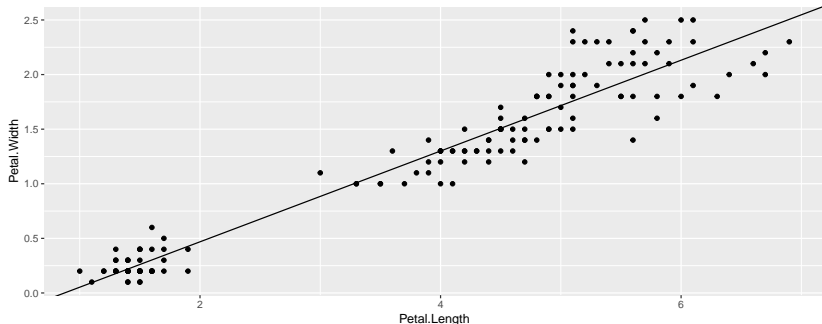
```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Petal.Width)) +  
  geom_point()
```





```
modelo <- lm(Petal.Width ~ Petal.Length, data=iris)
```

```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Petal.Width)) +  
  geom_point() +  
  geom_abline(intercept=coef(modelo)[1],  
              slope=coef(modelo)[2])
```





```
iris %>%  
  ggplot(aes(x=Petal.Length, y=Petal.Width)) +  
  geom_point() +  
  geom_smooth(method="lm", formula=y~x)
```

