Student reports

Laurent

16/03/2020

Data input

Add a first section called Data input, in which you will load the rWSBIM1207 package, use the interroA.csv() function to get the name of a csv file containing test results for a set of students, and read these data into R using read_csv. Display the few first observations and write a short sentence explaining the data.

```
## install.packages("BiocManager")
## install.packages("remotes")
## BiocManager::install("UCLouvain-CBIO/rWSBIM1207")
library(rWSBIM1207)
interroA.csv()
```

```
## [1] "/usr/local/lib/R/site-library/rWSBIM1207/extdata/interroA.csv"
```

```
library("tidyverse")
x <- read_csv(interroA.csv())
```

Attention, ne pas faire ceci!!!

```
## read_csv("/usr/local/lib/R/site-library/rWSBIM1207/extdata/interroA.csv")
```

ou

Affichage

```
x
```

```
## # A tibble: 100 x 8
##    id     height gender      X interro1 interro2 interro3 interro4
##    <chr>   <dbl> <chr>   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
##  1 A74890    168 M       1.43        16       18        7       10
##  2 A85494    167 M       1.05        15       18       13       NA
##  3 A51820    166 M       0.435        4       10       NA        7
##  4 A98669    164 M       0.715       15       15       18       13
##  5 A75521    171 M       0.917       18       10       17       NA
##  6 A96704    178 F      -2.66        11       20       14       17
##  7 A23214    155 M       1.11        12        2        8       14
##  8 A31124    177 M      -0.485       19        4        8       20
##  9 A80471    187 F       0.231       19       16       16        8
## 10 A21783    195 F      -0.295       13       11        8       20
## # ... with 90 more rows
```

Visualisation

Here, the goal is to visualise the score distributions for the four tests using ggplot2. These distributions will be visualised using boxplots. You will need to visualise these distribution for each test separately, and for male and female students.

As discussed during the course, we need data in a long format to be able to use ggplot2. Start by converting these data into a long format using pivot_longer() (or gather()). Display the first rows of these new data and write a short sentence describing them and the transformation you just applied.

```
xl <- x %>%
  pivot_longer(names_to = "interro",
               values_to = "res",
               starts_with("interro"))
```

```
x %>%
  gather(key = "interro",
         value = "res",
         starts_with("interro"))
```

```
## # A tibble: 400 x 6
##    id     height gender      X interro    res
##    <chr>   <dbl> <chr>   <dbl> <chr>    <dbl>
##  1 A74890    168 M       1.43  interro1    16
##  2 A85494    167 M       1.05  interro1    15
##  3 A51820    166 M       0.435 interro1     4
##  4 A98669    164 M       0.715 interro1    15
##  5 A75521    171 M       0.917 interro1    18
##  6 A96704    178 F      -2.66  interro1    11
##  7 A23214    155 M       1.11  interro1    12
##  8 A31124    177 M      -0.485 interro1    19
##  9 A80471    187 F       0.231 interro1    19
## 10 A21783    195 F      -0.295 interro1    13
## # ... with 390 more rows
```

```
x %>%
  pivot_longer(names_to = "interro",
               values_to = "res",
               5:8)
```

```
## # A tibble: 400 x 6
##    id     height gender     X interro    res
##    <chr>   <dbl> <chr>  <dbl> <chr>    <dbl>
##  1 A74890    168 M      1.43  interro1    16
##  2 A74890    168 M      1.43  interro2    18
##  3 A74890    168 M      1.43  interro3     7
##  4 A74890    168 M      1.43  interro4    10
##  5 A85494    167 M      1.05  interro1    15
##  6 A85494    167 M      1.05  interro2    18
##  7 A85494    167 M      1.05  interro3    13
##  8 A85494    167 M      1.05  interro4    NA
##  9 A51820    166 M      0.435 interro1     4
## 10 A51820    166 M      0.435 interro2    10
## # ... with 390 more rows
```

```
x %>%
  pivot_longer(names_to = "interro",
               values_to = "res",
               c(interro1, interro2,
                 interro3, interro4))
```

```
## # A tibble: 400 x 6
##    id     height gender     X interro    res
##    <chr>   <dbl> <chr>  <dbl> <chr>    <dbl>
##  1 A74890    168 M      1.43  interro1    16
##  2 A74890    168 M      1.43  interro2    18
##  3 A74890    168 M      1.43  interro3     7
##  4 A74890    168 M      1.43  interro4    10
##  5 A85494    167 M      1.05  interro1    15
##  6 A85494    167 M      1.05  interro2    18
##  7 A85494    167 M      1.05  interro3    13
##  8 A85494    167 M      1.05  interro4    NA
##  9 A51820    166 M      0.435 interro1     4
## 10 A51820    166 M      0.435 interro2    10
## # ... with 390 more rows
```

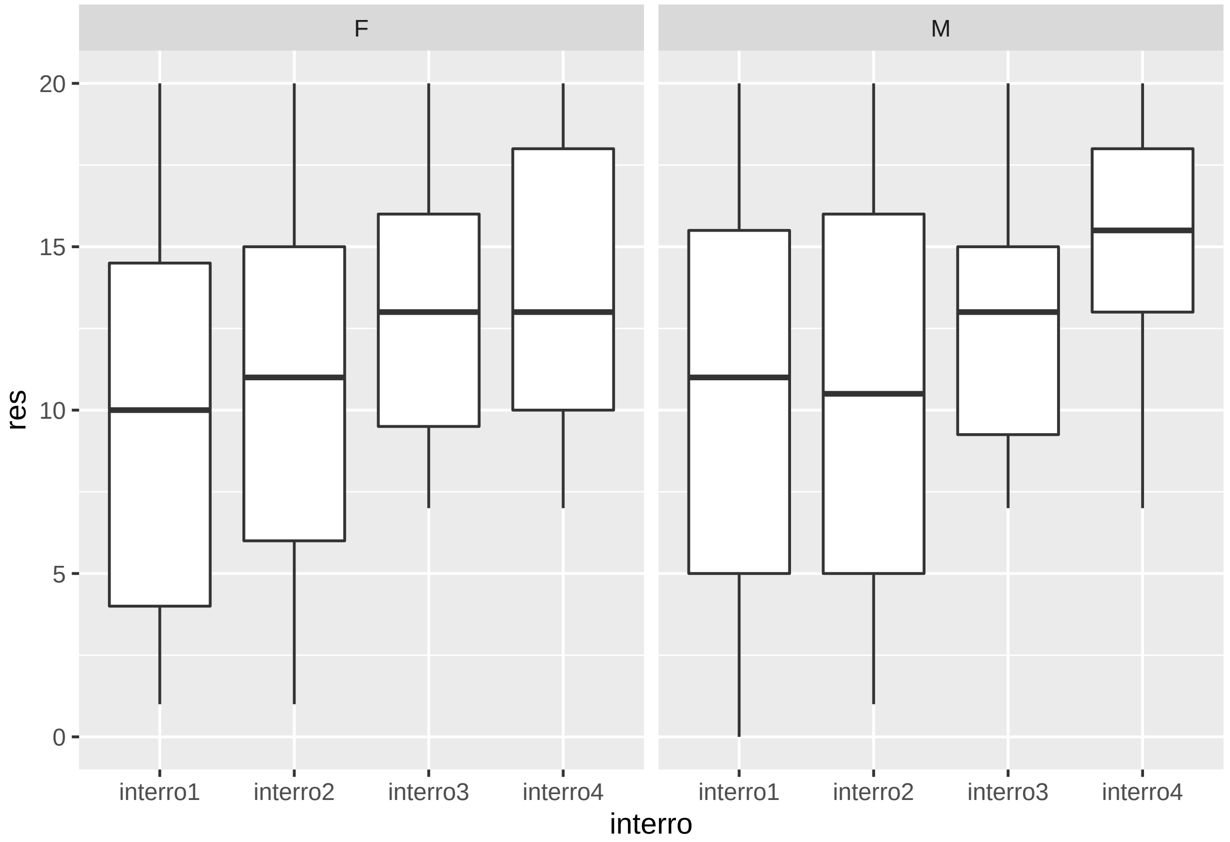
```
x %>%
  pivot_longer(names_to = "interro",
               values_to = "res",
               -(1:4))
```

```
## # A tibble: 400 x 6
##    id     height gender     X interro    res
##    <chr>   <dbl> <chr>  <dbl> <chr>    <dbl>
##  1 A74890    168 M      1.43  interro1    16
##  2 A74890    168 M      1.43  interro2    18
##  3 A74890    168 M      1.43  interro3     7
##  4 A74890    168 M      1.43  interro4    10
##  5 A85494    167 M      1.05  interro1    15
##  6 A85494    167 M      1.05  interro2    18
##  7 A85494    167 M      1.05  interro3    13
##  8 A85494    167 M      1.05  interro4    NA
##  9 A51820    166 M      0.435 interro1     4
## 10 A51820    166 M      0.435 interro2    10
## # ... with 390 more rows
```

Use ggplot2 to visualise the score distributions along boxplots for each test and for female and male students.

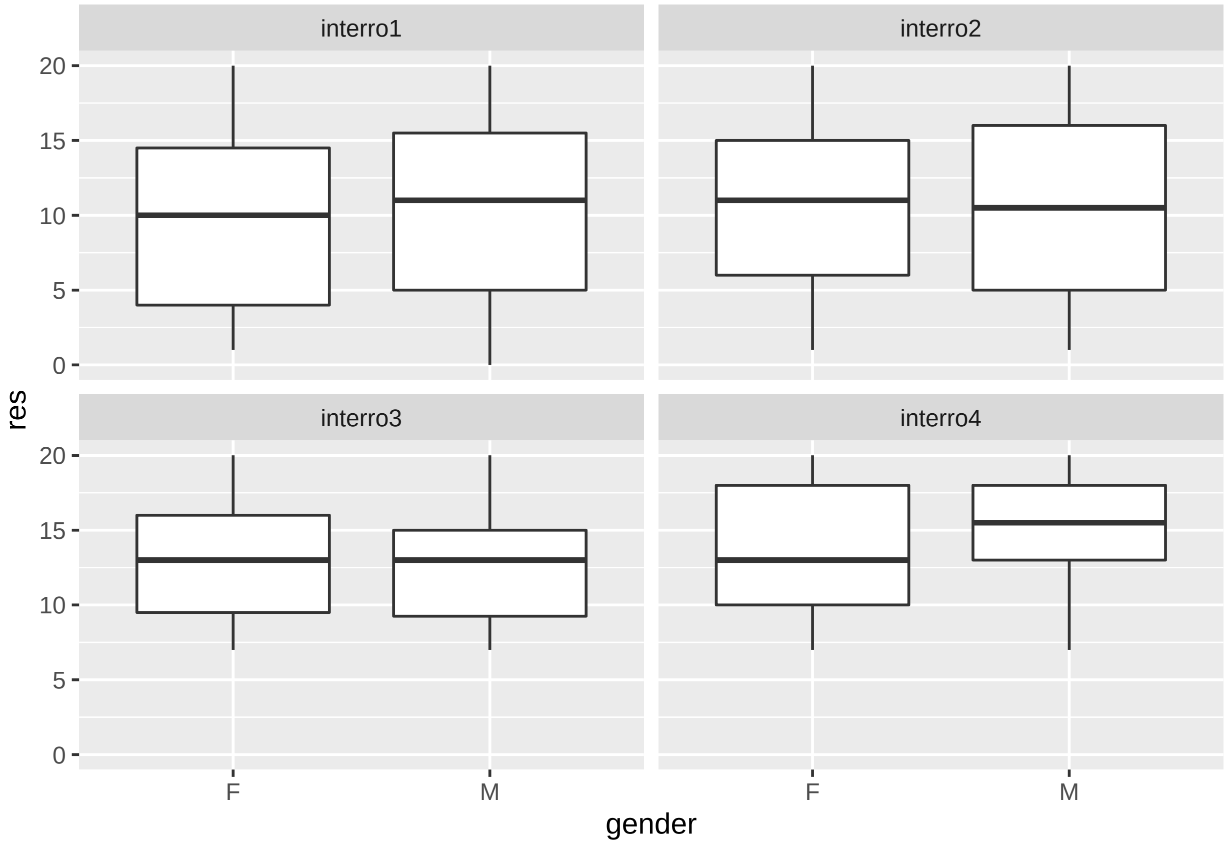
```
ggplot(xl, aes(x = interro, y = res)) +
  geom_boxplot() +
  facet_wrap(~ gender)
```

```
## Warning: Removed 18 rows containing non-finite values (stat_boxplot).
```



```
ggplot(xl, aes(x = gender, y = res)) +
  geom_boxplot() +
  facet_wrap(~ interro)
```

```
## Warning: Removed 18 rows containing non-finite values (stat_boxplot).
```


interro1
interro2
interro3
interro4
res
20
15
10
5
0
F
M
gender