scenario_2_r_markdown

Grace E. Seo

2/1/2022

Scenario 2

In Scenario 1, you performed some analysis and created a scatter plot of Pokemon bioinformatics class using student metadata.

To provide the result to your supervisor, you need to create a PDF output using R markdown.

Before you begin, let's learn how R markdown works briefly.

If you want to know more about R markdown, here is a great resource: https://bookdown.org/yihui/rmarkdown-cookbook/

In R markdown, you can create a Bookmark/table of contents by using #.

For example:

Header 1

Header 2

Header 3

Bullet points

If you want to include bullet points, use "-" in front of your sentence.

- Like this! Make sure you put a space before the dash!
- If you want to leave some blank lines between sentences, include a white space between two sentences by pressing "Enter".

Include a timestamp on this document

To know when this PDF file was generated, include a timestamp.

date: "Last compiled on 01 February, 2022, at 06:27:09"

Now, let's include all the codes from scenario 1 as well include outputs directly in this PDF file.

1. Load Pokemon class metadata

GitHub page: https://github.com/MMID-coding-workshop/2022-02-02-Introduction-to-R

Raw Pokemon class metadata: https://github.com/MMID-coding-workshop/2022-02-02-Introduction-to-R/blob/main/class_metadata.csv

```
## Raw csv file: https://github.com/MMID-coding-workshop/2022-02-02-Introduction-to-R/blob/main/class_m
## Two methods to load dataset: using a URL or directly from your folder.

# CODE: OBJECTNAME <- read.csv("FOLDERPATH/FILE.csv", header = TRUE)
# CODE: OBJECTNAME <- read.csv("URL/FILE.csv", header = TRUE)

#metadata <- read.csv("C:/Users/USERNAME/Desktop/intro_to_r/class_metadata.csv", header = TRUE)

metadata <- read.csv("https://raw.githubusercontent.com/MMID-coding-workshop/2022-02-02-Introduction-to

## NOTE: now you should see that "metadata" object is available in your "Environment" pane
```

View dataset

```
# CODE: View("OBJECTNAME")

View(metadata)

## NOTE: new window will pop-up beside this script. You can scroll through the content.

metadata # this will print data into console.
```

```
##    Number   Student_name Student_number Age Year_of_study        Phone
## 1       1            Mew        7991102  28             4 111-111-1111
## 2       2       Garchomp        7980475  18             5 111-111-1112
## 3       3 Kanto Starters        7779385  19             3 111-111-1113
## 4       4      Metagross        7620583  24             3 111-111-1114
## 5       5        Zoroark        8103955  25             3 111-111-1115
## 6       6      Tyranitar        8502931  21             4 111-111-1116
## 7       7           Eevee        8110291  29             4 111-111-1117
## 8       8        Snorlax        8110429  30             5 111-111-1118
## 9       9      Volcarona        7889281  34             4 111-111-1119
## 10     10         Mewtwo        8719237  21             3 111-111-1120
## 11     11      Charizard        7817241  22             3 111-111-1121
## 12     12        Milotic        7992930  21             3 111-111-1122
## 13     13         Kyurem        7293148  24             3 111-111-1123
## 14     14        Haxorus        7920394  31             3 111-111-1124
## 15     15         Arceus        7329523  19             4 111-111-1125
## 16     16      Ninetales        7978149  21             4 111-111-1126
## 17     17         Espeon        7918938  24             4 111-111-1127
## 18     18         Golurk        7083872  23             4 111-111-1128
## 19     19         Lapras        7051934  22             5 111-111-1129
## 20     20        Pikachu        7094583  24             5 111-111-1130
##                Address    Address_city               Major
## 1  111 Pokemon Street     Pallet Town             Biology
```

```
## 2  112 Pokemon Street     Pallet Town              Biology
## 3  113 Pokemon Street     Pallet Town         Microbiology
## 4  114 Pokemon Street   Viridian City         Microbiology
## 5  115 Pokemon Street     Pewter City         Microbiology
## 6  116 Pokemon Street   Viridian City Medical Microbiology
## 7  117 Pokemon Street   Viridian City          Mathematics
## 8  118 Pokemon Street     Pewter City          Mathematics
## 9  119 Pokemon Street   Lavender Town           Statistics
## 10 120 Pokemon Street   Lavender Town Medical Microbiology
## 11 121 Pokemon Street   Lavender Town Medical Microbiology
## 12 122 Pokemon Street    Celadon City Medical Microbiology
## 13 123 Pokemon Street    Celadon City Medical Microbiology
## 14 124 Pokemon Street    Celadon City       Bioinformatics
## 15 125 Pokemon Street    Saffron City       Bioinformatics
## 16 126 Pokemon Street    Saffron City           Statistics
## 17 127 Pokemon Street   Cerulean City       Bioinformatics
## 18 128 Pokemon Street   Cerulean City            Chemistry
## 19 129 Pokemon Street Vermillion City         Biochemistry
## 20 130 Pokemon Street Vermillion City         Biochemistry
```

2. Install and load necessary package for your analysis if not already installed

```
## For simple analyses, we will install tidyverse package available in CRAN
## NOTE: tidyverse package includes tidyr, dplyr, ggplot2 and many more!
# CODE: install.packages("PACKAGENAME")

#install.packages("tidyverse")

## NOTE: if you already have tidyverse installed, comment out the above code so you don't repeat instal


## Load installed package == library
# CODE: library(PACKAGENAME)

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

3. Determine how many students are in your class

HINT: Row number matches student number

```
#CODE: nrow(OBJECTNAME) # counts how many rows are present in a dataframe

nrow(metadata)
```

```
## [1] 20
```

ANSWER: 20 students

4. Determine how many columns are present in your metadata

```
# CODE: ncol(OBJECTNAME) # counts how many columns are present in a dataframe

ncol(metadata)
```

```
## [1] 9
```

ANSWER: 9

5. Show the list of column headers in the metadata

```
# CODE: colnames(OBJECTNAME) # prints out list of column headers

colnames(metadata)
```

```
## [1] "Number"         "Student_name"   "Student_number" "Age"
## [5] "Year_of_study"  "Phone"          "Address"        "Address_city"
## [9] "Major"
```

ANSWER: [1] “Number” “Student_name” “Student_number” “Age” “Year_of_study” “Phone”
[7] “Address” “Address_city” “Major”

6. Print out row contents of only column "Student_name"

```
# CODE: OBJECTNAME$COLNAME

metadata$Student_name
```

```
##  [1] "Mew"            "Garchomp"       "Kanto Starters" "Metagross"
##  [5] "Zoroark"        "Tyranitar"      "Eevee"          "Snorlax"
##  [9] "Volcarona"      "Mewtwo"         "Charizard"      "Milotic"
## [13] "Kyurem"         "Haxorus"        "Arceus"         "Ninetales"
## [17] "Espeon"         "Golurk"         "Lapras"         "Pikachu"
```

ANSWER: [1] "Mew" "Garchomp" "Kanto Starters" "Metagross" "Zoroark" "Tyranitar"
[7] "Eevee" "Snorlax" "Volcarona" "Mewtwo" "Charizard" "Milotic"
[13] "Kyurem" "Haxorus" "Arceus" "Ninetales" "Espeon" "Golurk"
[19] "Lapras" "Pikachu"

7. Determine average age of all students in this class

```
# CODE: mean(OBJECTNAME$COLNAME)

mean(metadata$Age)
```

```
## [1] 24
```

ANSWER: 24

8. Determine the Pikachu's student number

```
# CODE: OBJECTNAME$COLNAME_CONTAIN_VALUE_YOU_WANT[OBJECTNAME$COLNAME_YOUR_SEARCH == "NAME"]

metadata$Student_number[metadata$Student_name=="Pikachu"]
```

```
## [1] 7094583
```

ANSWER: 7094583

9. Scatter plot by student's Year_of_study vs. Age and create a figure legend by student's study Major

```
# CODE: NEWOBJECT <- ggplot(data = OBJECTNAME, aes(x = X_VALUE, y = Y_VALUE, colour = GROUP)) +
#                    geom_point()

# CODE: NEWOBJECT # to view the content


plot <- ggplot(data = metadata, aes(x = Year_of_study, y = Age, colour = Major)) +
        geom_point()

plot # view the plot
```

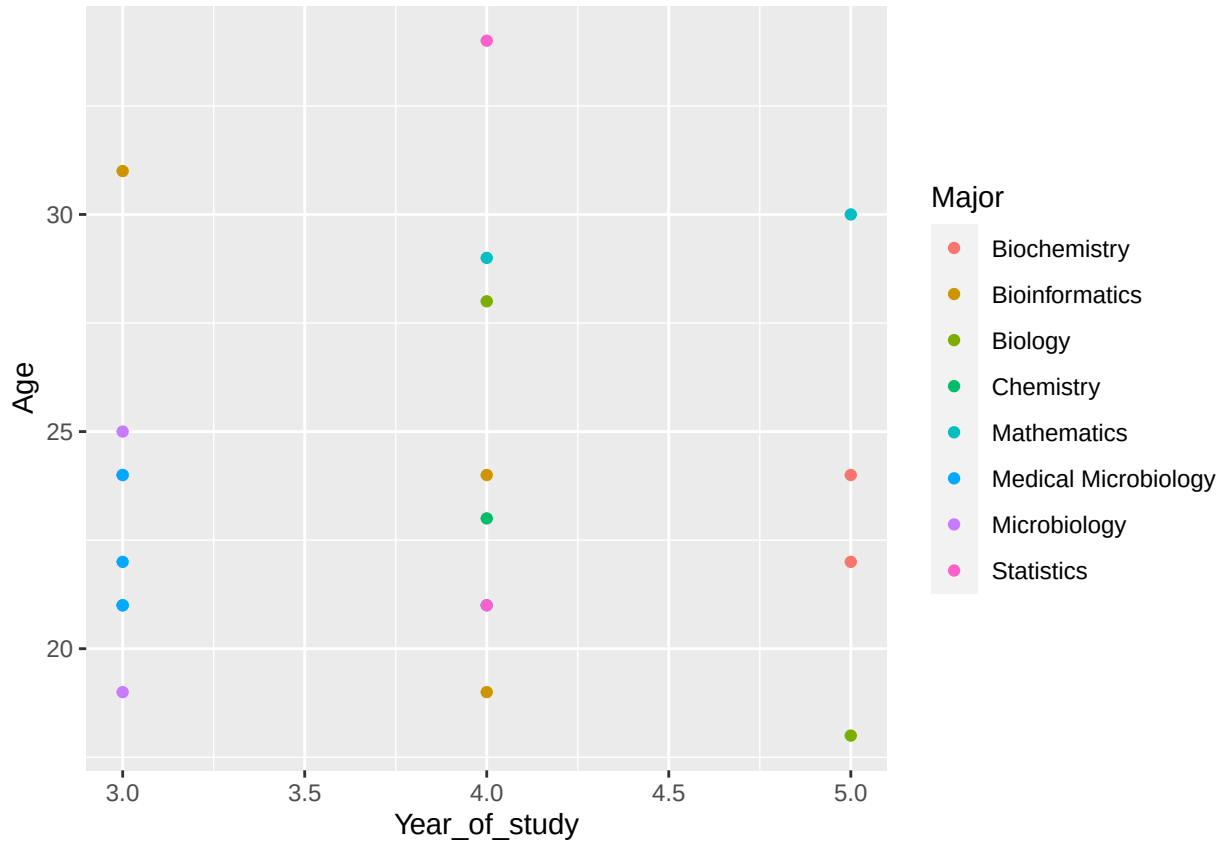

Figure 1: Student's Year of Study by Age and Major represented as simple scatter plot.

10. Scatter plot is too small to see.

Increase the size of each scatter point, remove grey background and rename the x-axis to "Year of Study".

```
plot_2 <- ggplot(data = metadata, aes(x = Year_of_study, y = Age, colour = Major)) +
    geom_point(size = 5) +
    xlab("Year of Study") +
    theme_bw()

plot_2 # view the plot
```

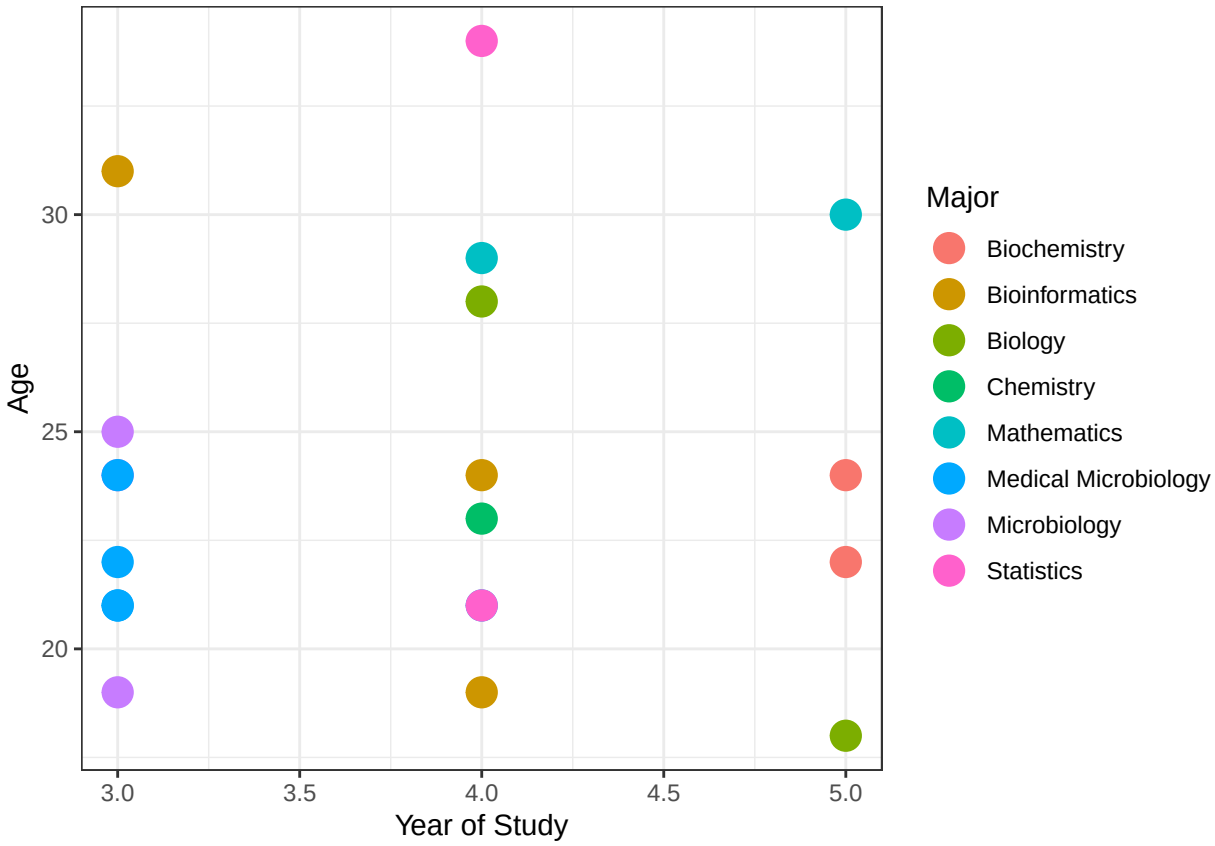

Figure 2: Student's Year of Study by Age and Major represented as scatter plot.

This is it! Now click "Knit" at top on this R markdown –> "Knit to PDF" or "Knit to HTML" and you'll get a file output.

The end of scenario 2

Upcoming workshops

Now try this with your own code and plots. In upcoming R workshops, you'll be able to learn in-depth on how to do:

- more sophisticated data transformation
- include more options and make fancier graphs in ggplot, ggtree, etc.
- perform RNA-seq data analyses
- and build machine learning models!

References

If you want to practice R markdown in detail, here are helpful resources:

1. Yihui Xie, Christophe Dervieux, Emily Riederer. 2022. R Markdown Cookbook. Available at https://bookdown.org/yihui/rmarkdown-cookbook/.
2. Chester Ismay and Patrick C. Kennedy. 2021. Getting Used to R, RStudio, and R Markdown. https://ismayc.github.io/rbasics-book/index.html.