rbtl - Data Visualisation

Lars Schöbitz

Global Health Engineering - ETH Zurich

2022-05-04

Today

1. Research Project Report
2. Solving coding problems
3. Working collaboratively with git
    - Live Coding Exercise
4. Exploratory Data Analysis with `ggplot2`
    - Live Coding Exercise
    - Programming Exercise
5. Homework Assignment 11

Learning Objectives

1. Learners can describe the four main aesthetic mappings that can be used to visualise data using the ggplot2 R Package
2. Learners can control the colour scaling applied to a plot using colour as an aesthetic mapping
3. Learners can compare three different geoms and their use case
4. Learners can apply a theme to control font types and sizes within a plot

Research project report

GitHub issues / Slack

[TODO: List of aggregated questions and answers]

Solving coding problems

Tipps for search engines

- Use actionable verbs that describe what you want to do
- Be specific
- Add R to the search query
- Add the name of the R package name to the search query
- Scroll through the top 5 results (don’t just pick the first)

Example: “How to remove a legend from a plot in R ggplot2”

Stack Overflow

What is it?

- The biggest support network for (coding) problems
- Can be initimidating at first
- Upvote system

Workflow

- First, briefly read the question that was posted
- Then, read the answer marked as “correct”
- Then, read one or two more answers with high votes
- Then, check out the “Linked” posts
- Always give credit for the solution

Give credit

528

from r cookbook, where bp is your ggplot:

Remove legend for a particular aesthetic (fill):

```
bp + guides(fill="none")
```

It can also be done when specifying the scale:

```
bp + scale_fill_discrete(guide="none")
```

This removes all legends:

```
bp + theme(legend.position="none")
```

Share Edit Follow Flag

edited Dec 2, 2021 at 7:07

Andrew Morris

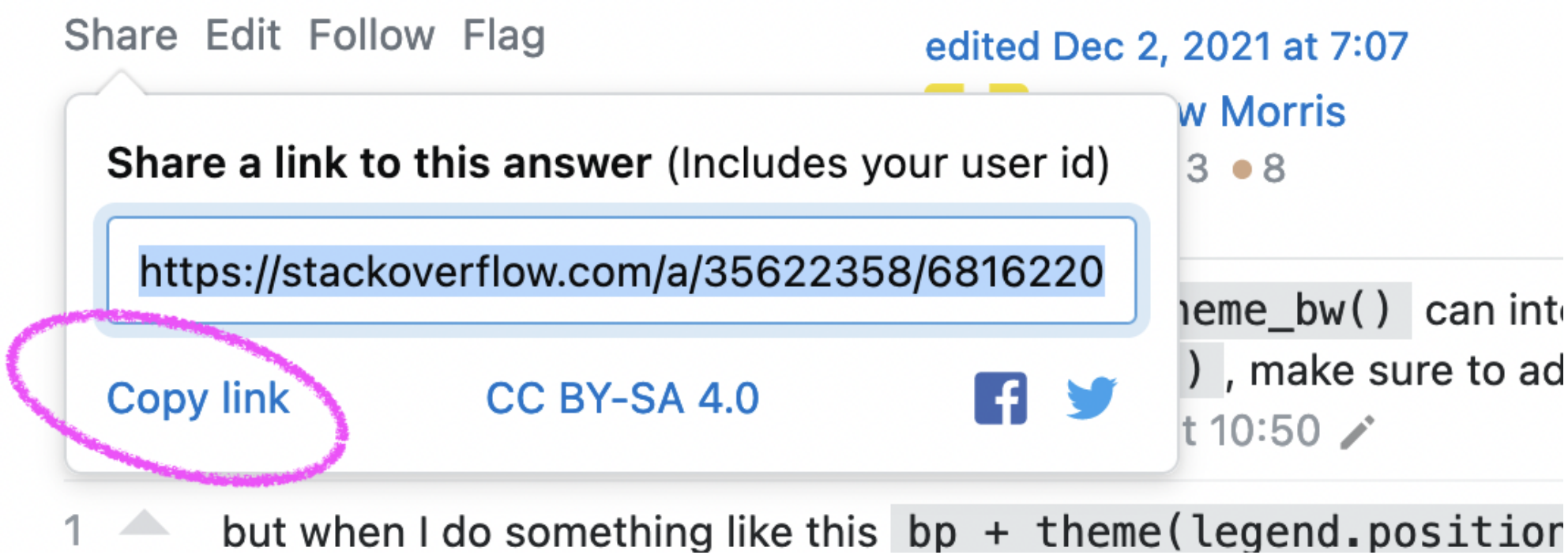
408 ● 3 ● 8

answered Feb 25, 2016 at 8:48

user3490026

5,388 ● 1 ● 9 ● 4

Give credit


Share Edit Follow Flag
edited Dec 2, 2021 at 7:07
w Morris
3 8
Share a link to this answer (Includes your user id)
https://stackoverflow.com/a/35622358/6816220
Copy link
CC BY-SA 4.0
neme_bw() can int
) , make sure to ad
t 10:50
1
but when I do something like this bp + theme(legend.positior

Give credit

```
ggplot(data = global_waste_data_kg_year,
       mapping = aes(x = income_id,
                     y = capita_kg_year,
                     color = income_id)) +
  ## Remove legend ref: https://stackoverflow.com/a/35622358/6816220
  theme(legend.position = "none")
```

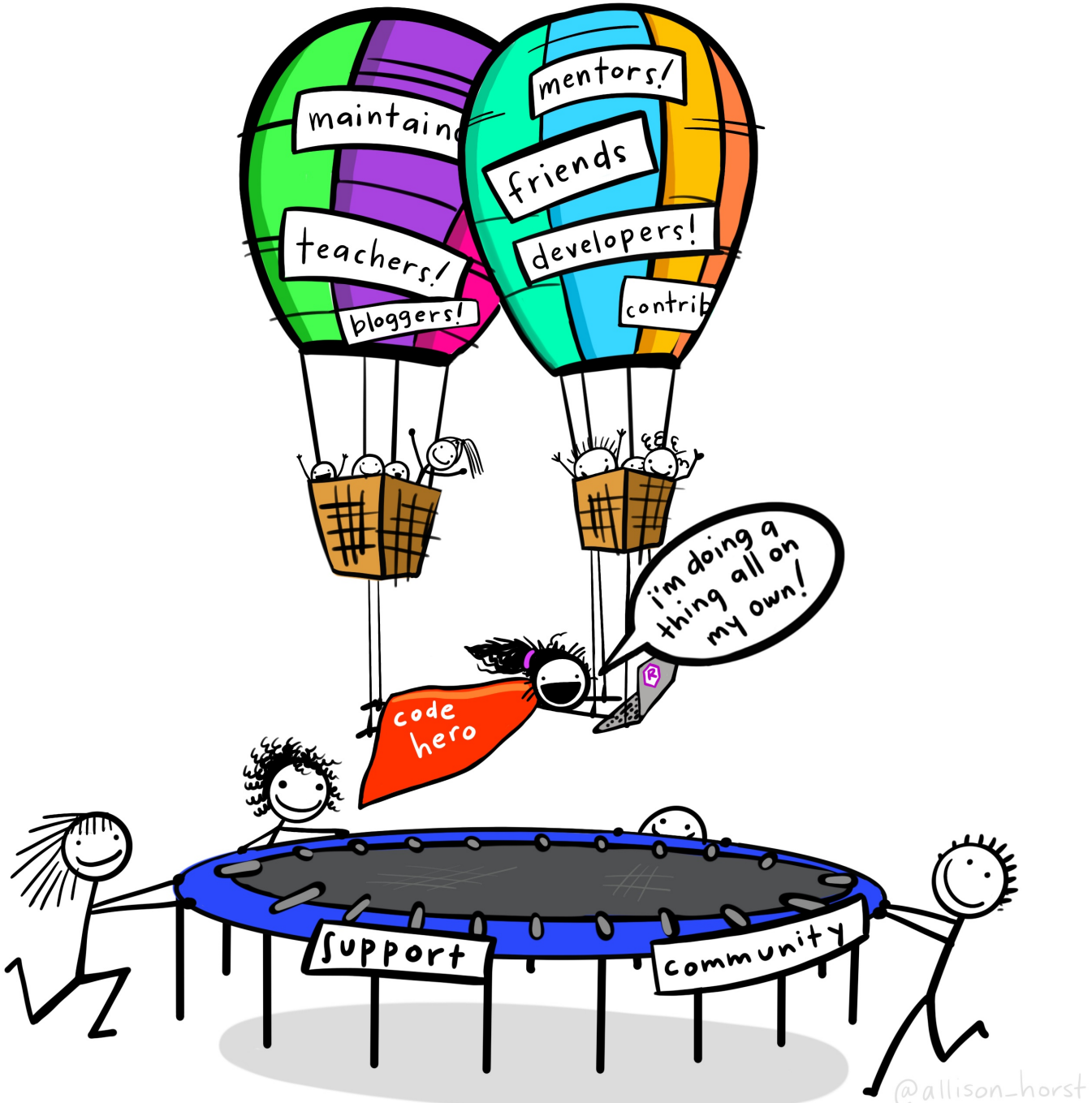
Other sources for help

- RStudio Community Forum: https://community.rstudio.com/
- Our rbtl Slack channel
- Documentation websites: https://dplyr.tidyverse.org/
- Twitter community: #rstats

Artwork by @allison_horst

Minimal reproducible example (reprex)

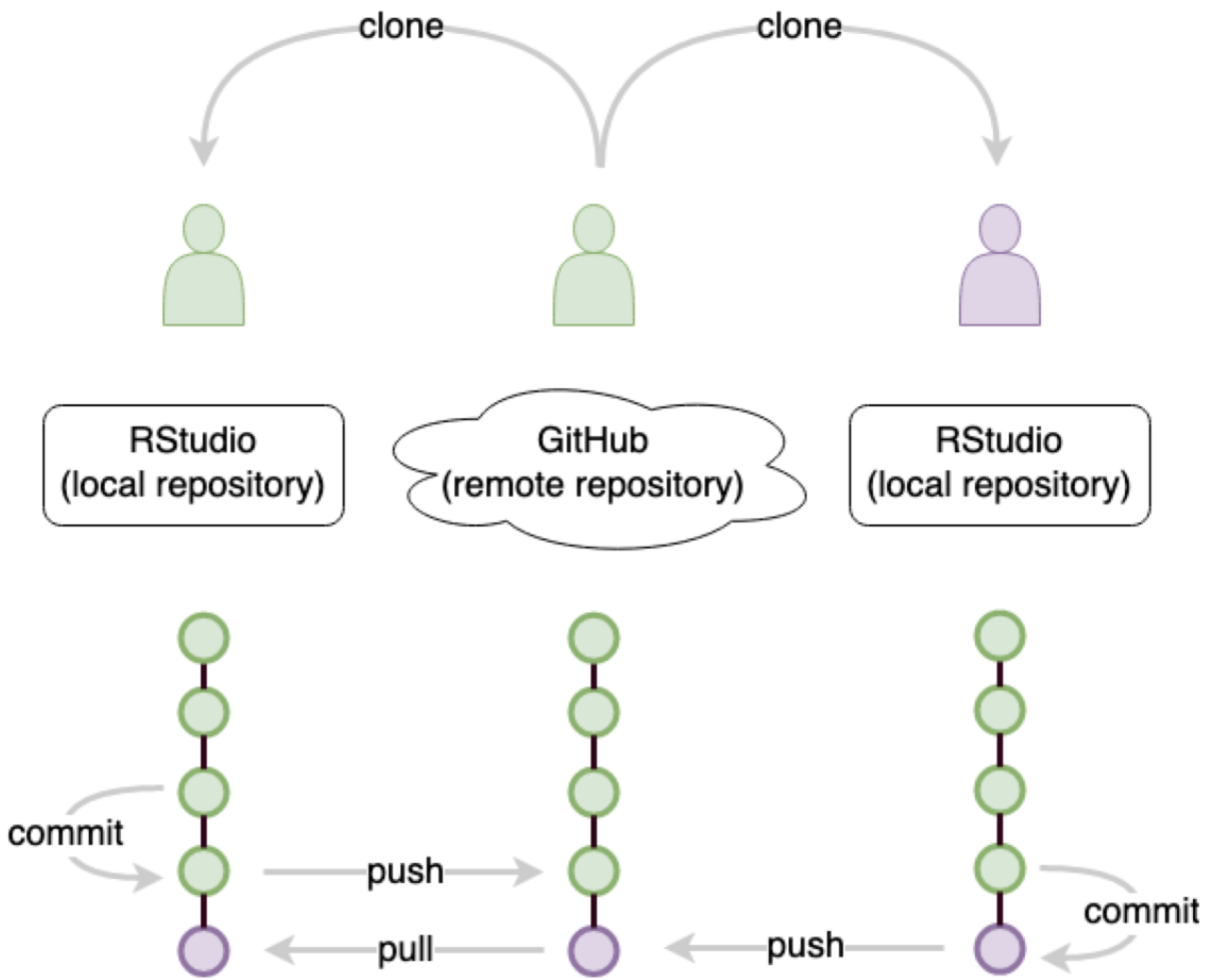
- Needed when asking questions online
- We will practice this in another class
- Good support information: https://www.tidyverse.org/help/#reprex



Artwork by @allison_horst

Working collaboratively with git

pull first, and push often



Live Coding Exercise

1. Open the repo for your team project report on RStudio Cloud
2. Open the file: `01-introduction.qmd`
3. Use your Sticky Notes to let me know when you are ready.

Git help

You can find the merge conflict workflow documented in our git help document for the course:

rbtl-fs22/git-help

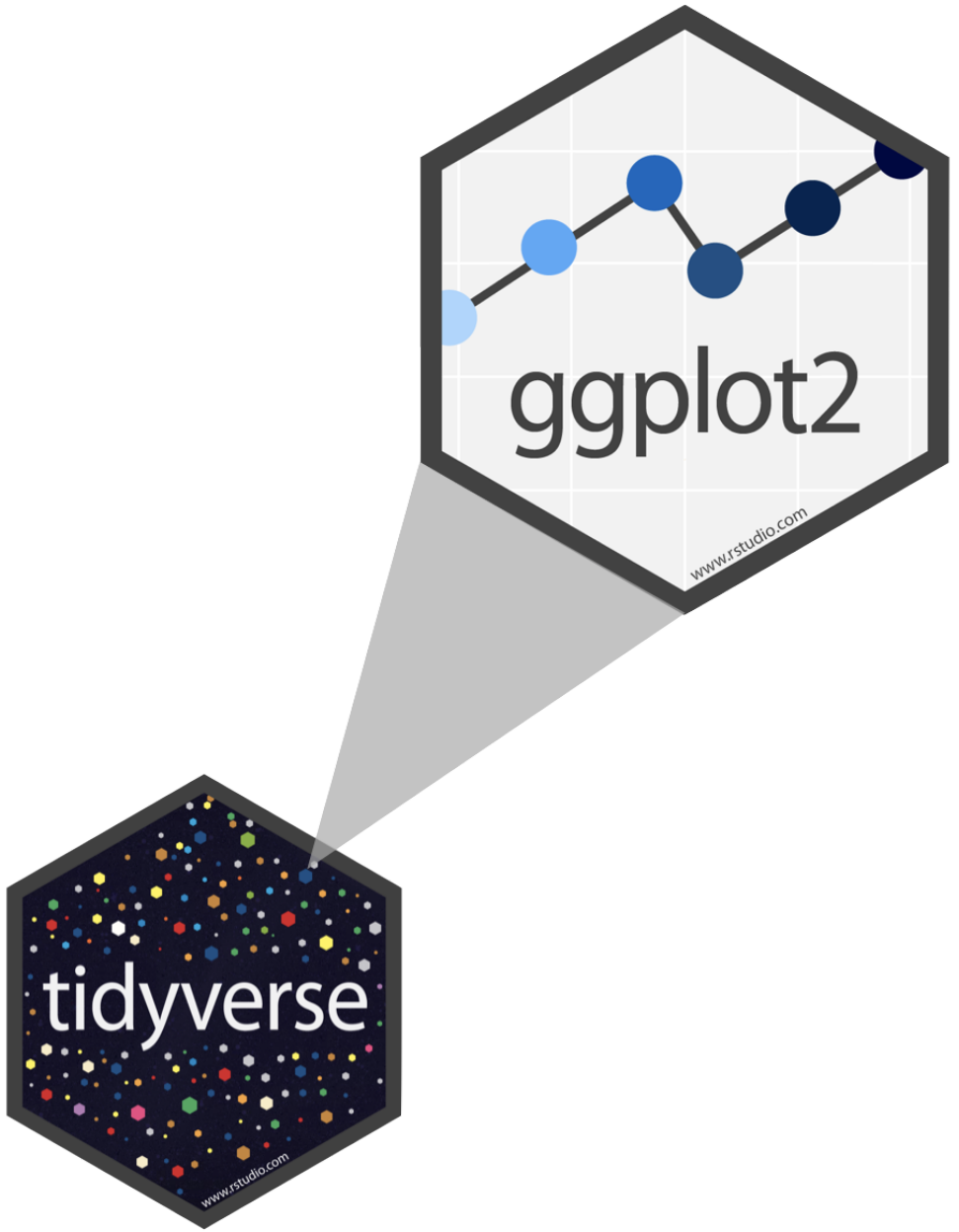
The best online resource for working with git is:

Happy Git and GitHub for the useR by Jenny Bryan

Exploratory Data Analysis with `ggplot2`

R Package `ggplot2`

- **ggplot2** is tidyverse's data visualization package
- `gg` in `ggplot2` stands for Grammar of Graphics
- Inspired by the book **Grammar of Graphics** by Leland Wilkinson
- **Documentation:** https://ggplot2.tidyverse.org/
- **Book**: https://ggplot2-book.org



Slide adapted from datasciencebox.org

Code structure

- `ggplot()` is the main function in ggplot2
- Plots are constructed in layers
- Structure of the code for plots can be summarized as

```
ggplot(data = [dataset],
       mapping = aes(x = [x-variable],
                     y = [y-variable])) +
   geom_xxx() +
   other options
```

Code structure

```
1  ggplot()
```

Code structure

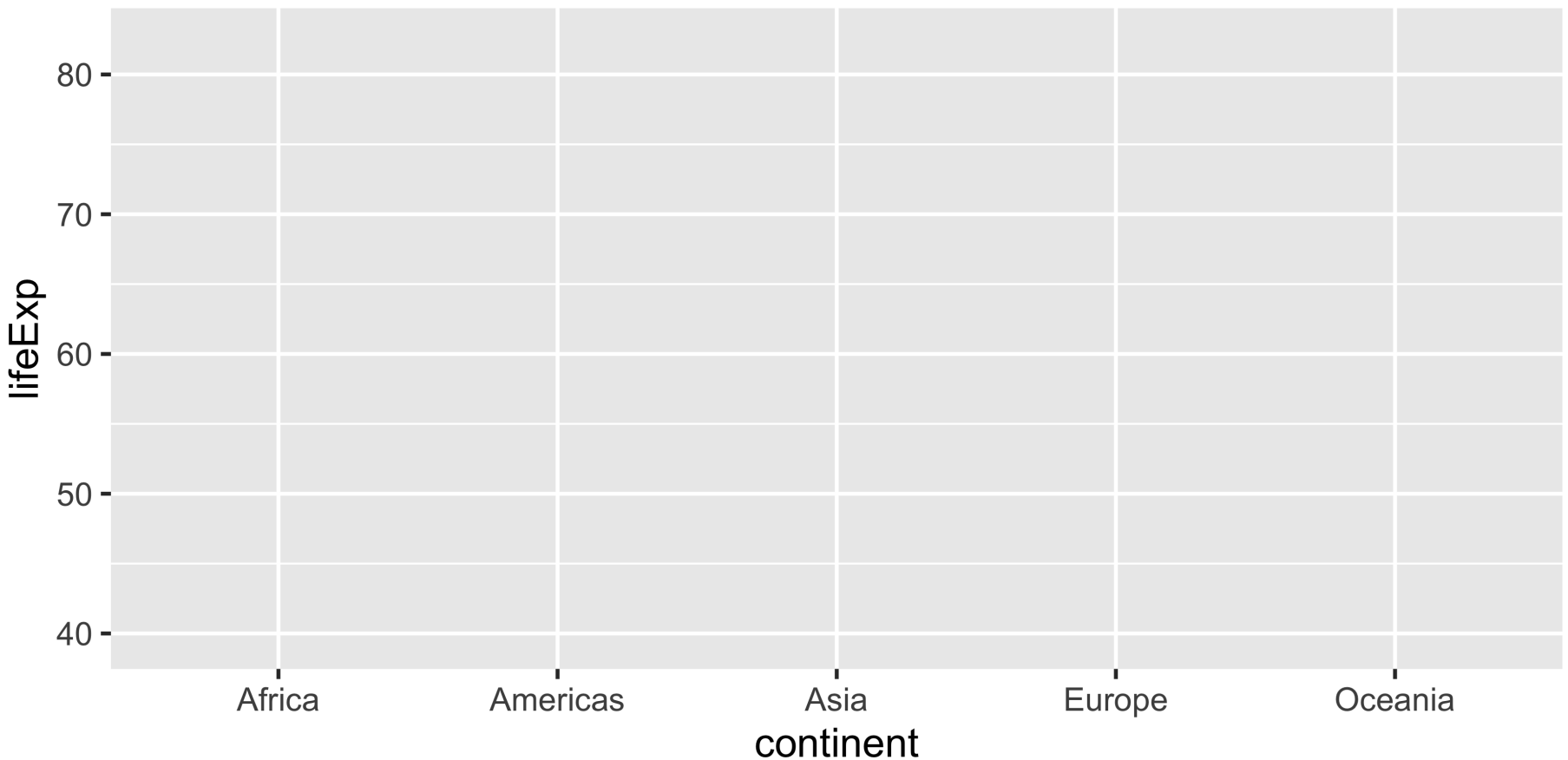
```
1  ggplot(data = gapminder_yr_2007)
```

Code structure

```
ggplot(data = gapminder_yr_2007,
       mapping = aes())
```

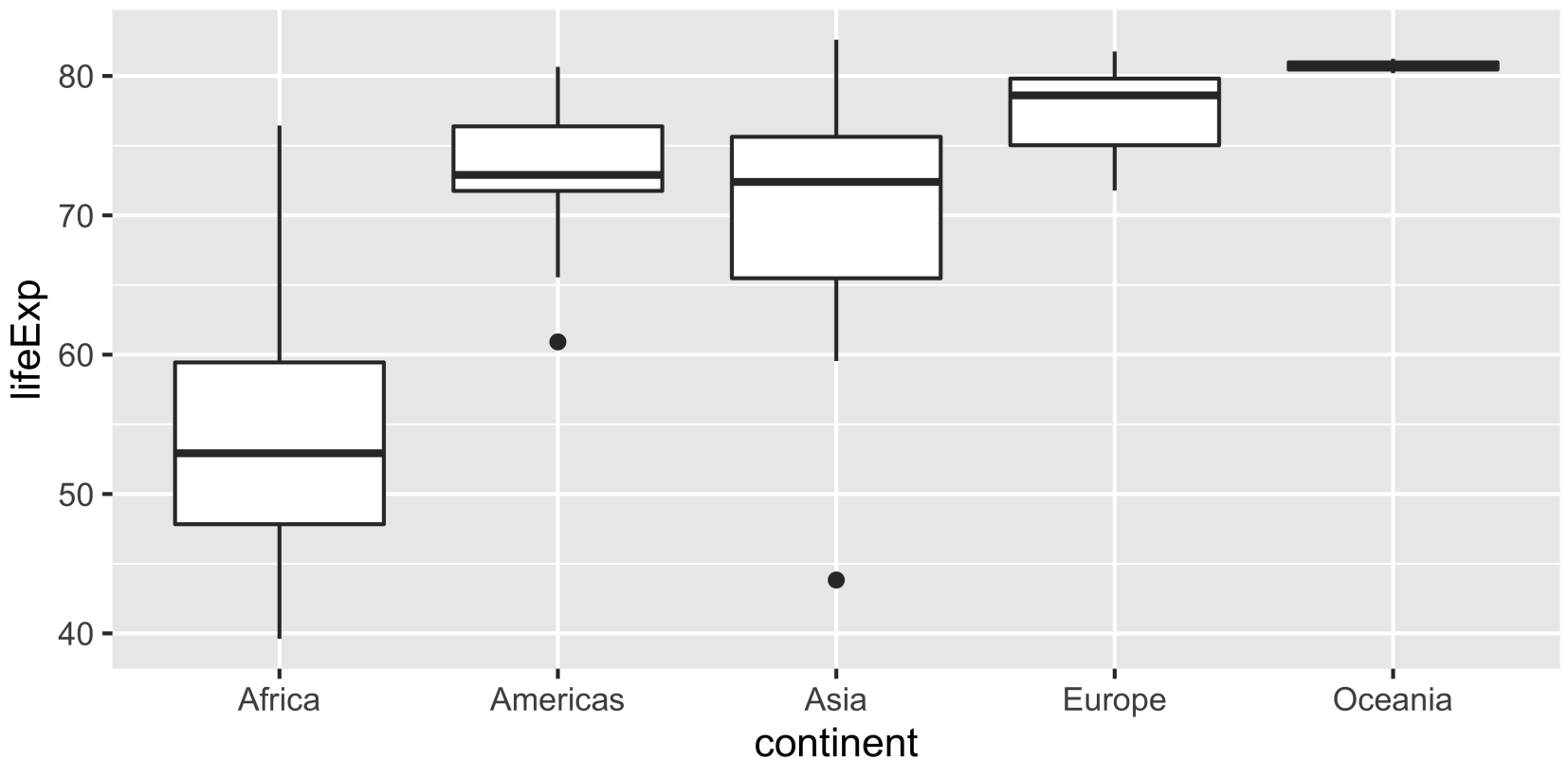
Code structure

```
1  ggplot(data = gapminder_yr_2007,
2         mapping = aes(x = continent,
3                       y = lifeExp))
```



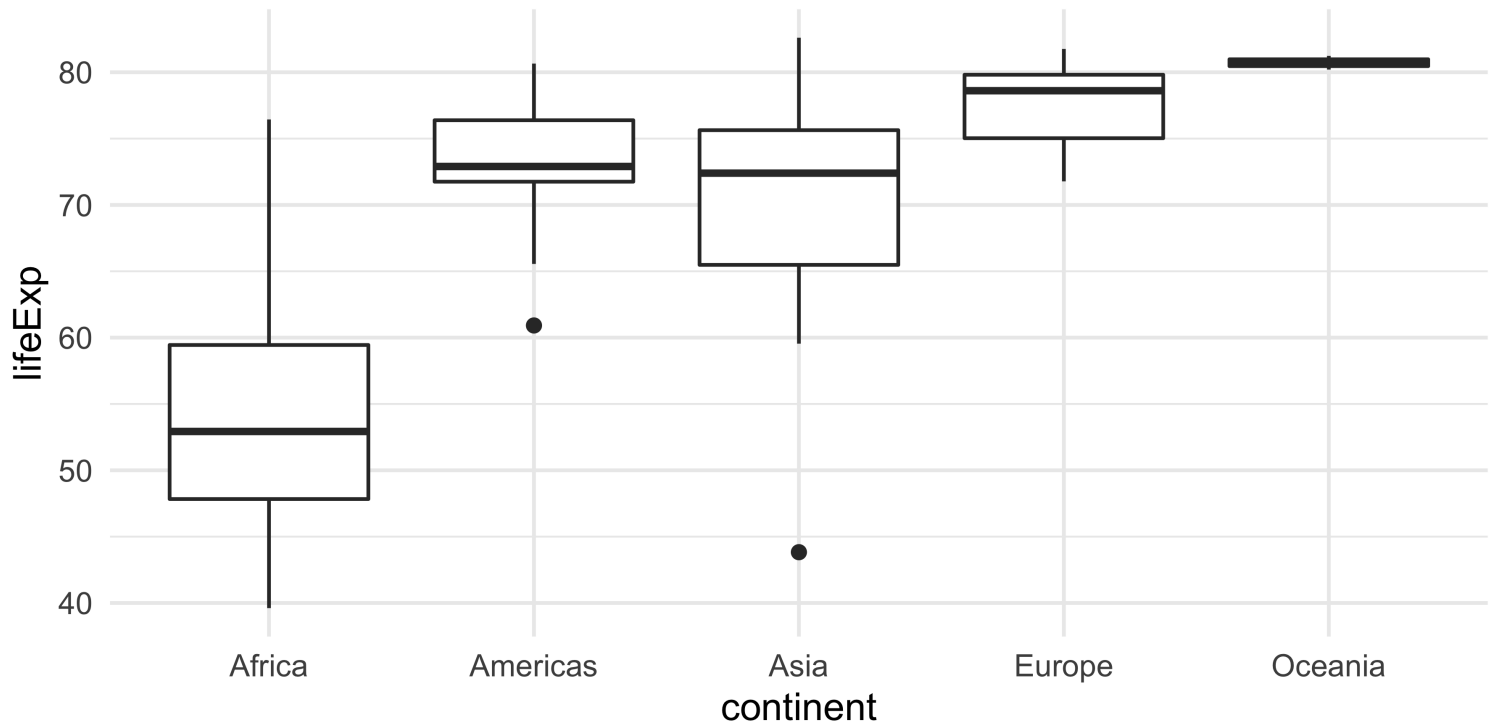
Code structure

```
1  ggplot(data = gapminder_yr_2007,
2         mapping = aes(x = continent,
3                       y = lifeExp)) +
4    geom_boxplot()
```



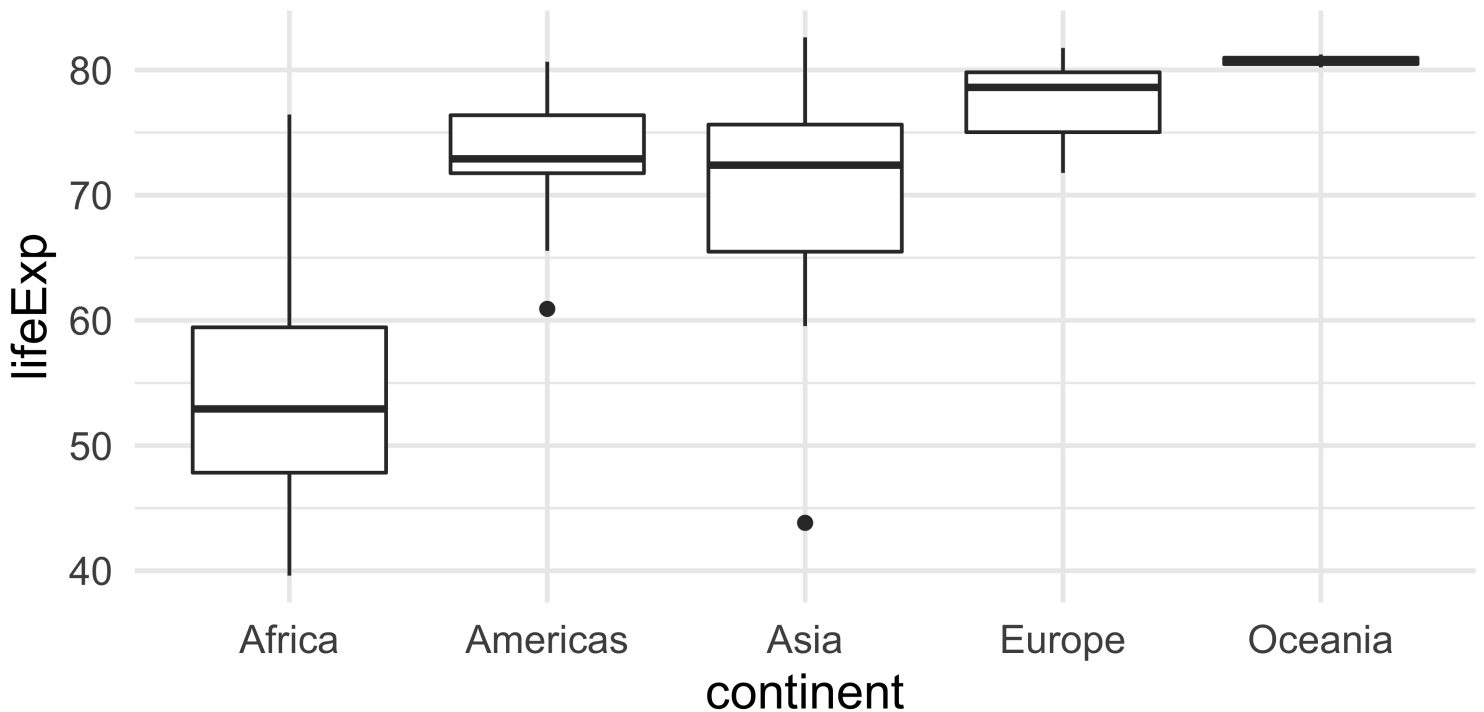
Code structure

```
1  ggplot(data = gapminder_yr_2007,
2         mapping = aes(x = continent,
3                       y = lifeExp)) +
4    geom_boxplot() +
5    theme_minimal()
```



Code structure

```
1  ggplot(data = gapminder_yr_2007,
2         mapping = aes(x = continent,
3                       y = lifeExp)) +
4    geom_boxplot() +
5    theme_minimal(base_size = 14)
```



Break

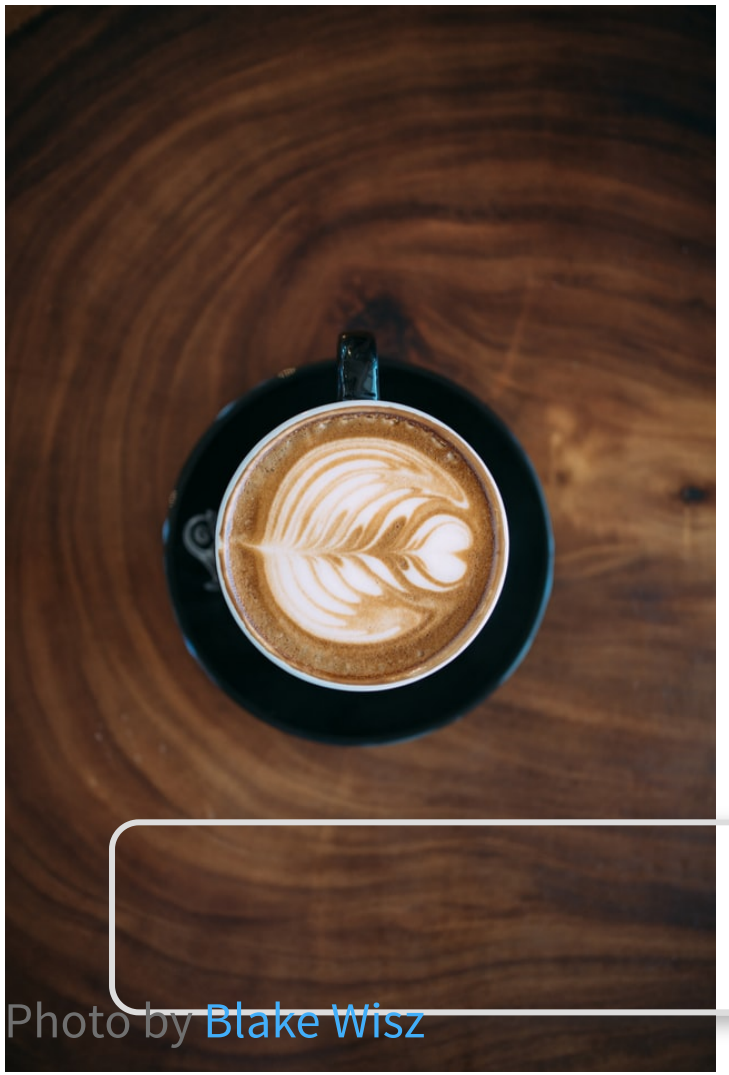


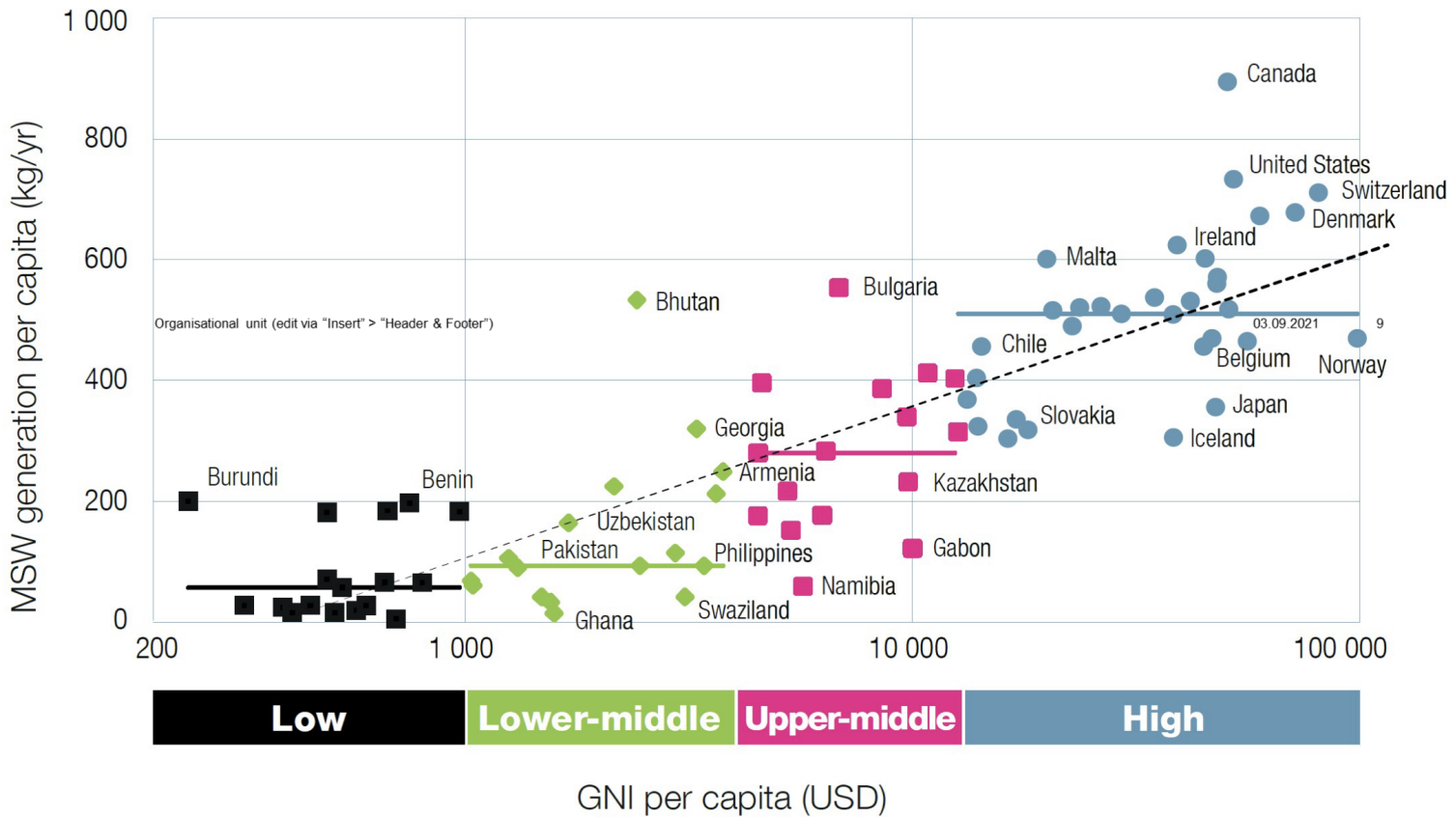
Photo by Blake Wisz

15:00

Live Coding Exercise - Goal

Generation

Switzerland has one of the highest municipal solid waste volumes in the world, at 716 kg of waste per person and year. *Nearly 53% of it is recycled.*



Live Coding Exercise

ae-11-data-science-lifecycle

1. Head over to the GitHub Organisation for the course.
2. Find the repo for week 11 that has your GitHub username.
3. Clone the repo with your username to the RStudio Cloud.
4. Open the file: `ae-11a-data-visualisation.qmd`
5. Use your Sticky Notes to let me know when you are ready.

Break

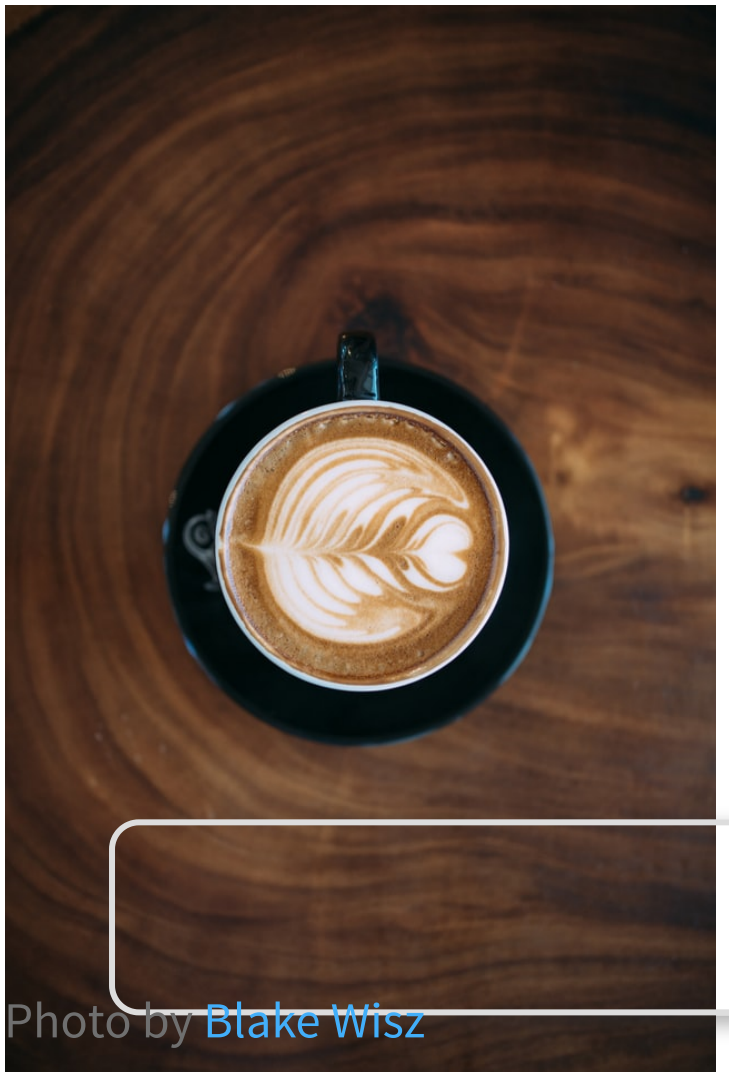
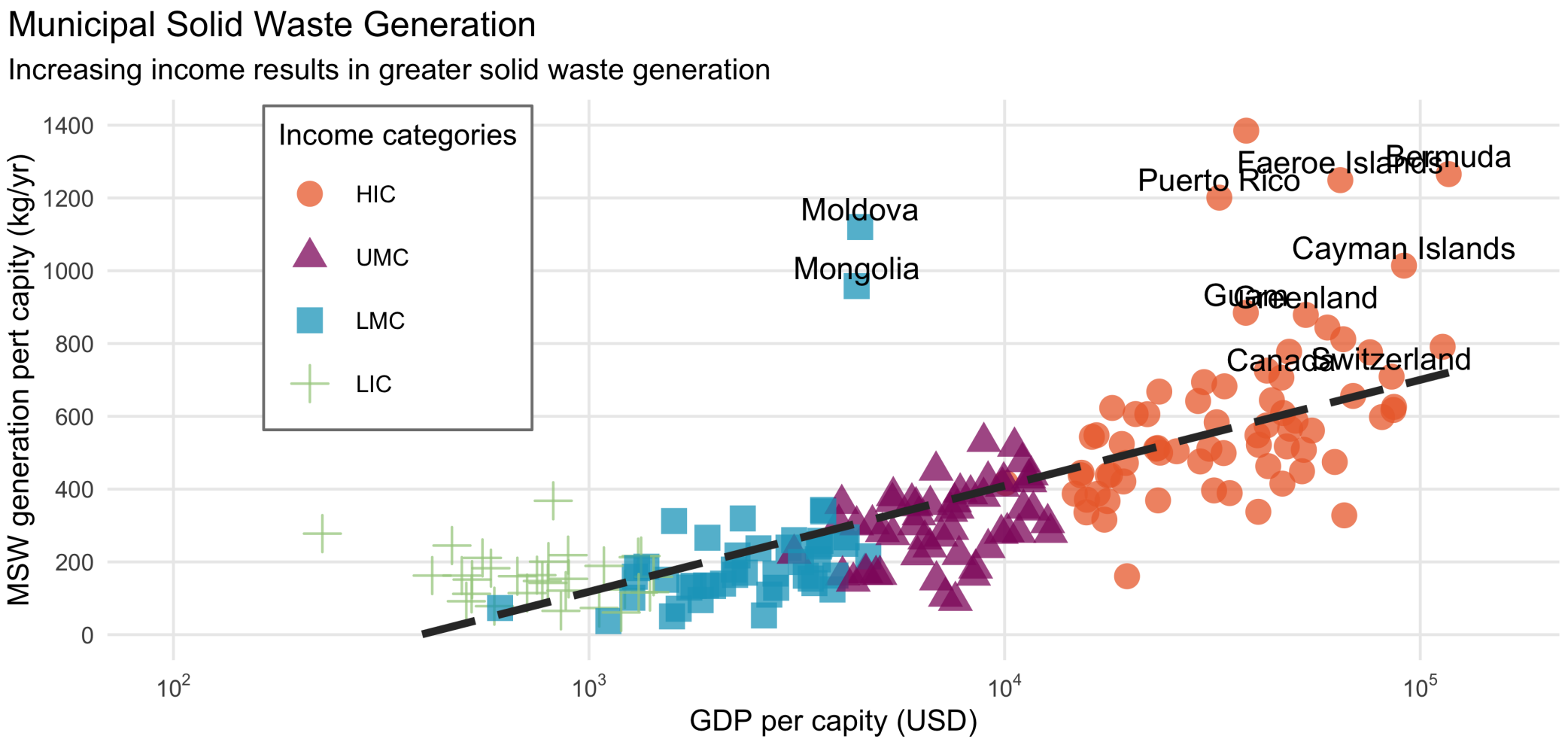


Photo by Blake Wisz

10:00

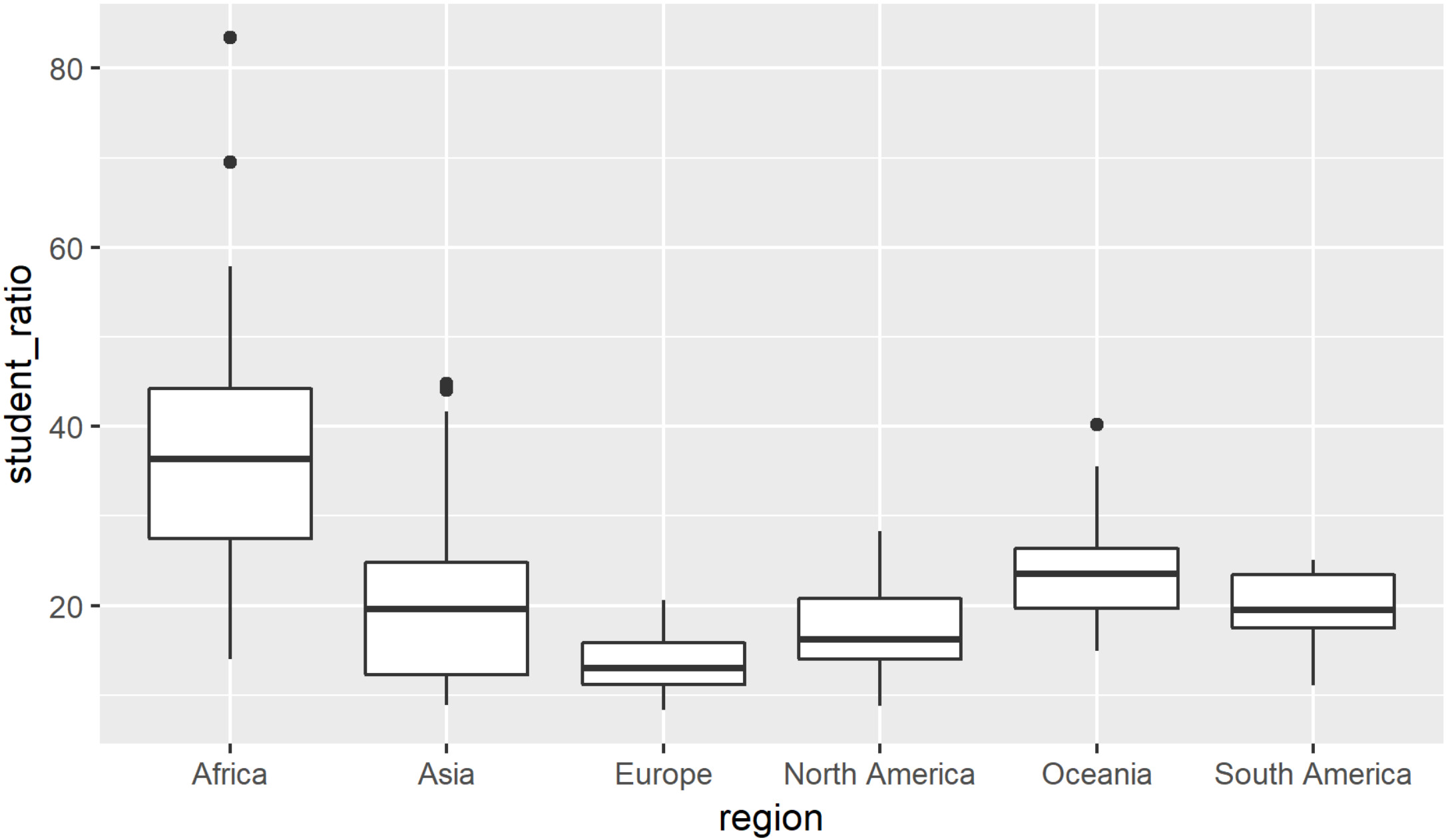
Live Coding Exercise - Result



Visualising numerical data

Types of variables

The Evolution of a ggplot



Data: UNESCO Institute for Statistics
Visualization by Cédric Scherer

Tutorial: https://www.cedricscherer.com/2019/05/17/the-evolution-of-a-ggplot-ep.-1/

data-to-viz.com

Homework Assignment

Submission

- All details in assignment week 11
- Due: Wednesday, 12th May at 23:59 (2 points)

Evaluation

- 5 mins
- anonymous
- after each lecture

https://forms.gle/HbCPbG9Yb7iDJ2jW6

Programming

ae-11-data-visualisation

1. Open the file: `ae-11b-data-visualisation.qmd`
2. Work through the exercises
3. Use your sticky notes to indicate if you need support

30:00

Thanks! 

Slides created via revealjs and Quarto: https://quarto.org/docs/presentations/revealjs/ Access slides as PDF on GitHub