Graded Assignment

Data Visualization 2024-2025

Information and instructions:

Assignment rules:

- You have 2 hours to finish this assignment. Students eligible for extra time receive an additional 20 minutes (10 minutes per hour). Distribute your time wisely, we recommend 50:50 for Task 1 and Task 2.
- You can use any resources such as course material, internet, StackOverflow, **except**:
  - help from anyone else, including posting questions on, for example, a forum,
  - help from ChatGPT or any other AI system that can give you (most) of the code.
- You may only use one screen, so you cannot use a tablet or any extra extra screen during the exam.

Invigilating

To check whether you abide by the above rules, there will be two invigilating systems:

- In-person invigilators: these persons walk around to check your screens.
- Invigilating via Zoom: we will monitor you screens via Zoom (see detailed instructions below).

If we notice that you are using materials or a system/website that you are not allowed to use, or that you use any way of communicating with others, this will be reported to the Board of Examiners and they will decide whether your assignment is valid.

Zoom instructions:

- Open the Zoom meeting, either in your Zoom app or in the Browser:
  - click on "link to the meeting"
  - or go to your Zoom application or https://www.zoom.us/join and join the meeting with the following details:
    - Meeting ID: 794 341 3633
    - Passcode: Xn#VR6je
- Use as your screen name your full name~~, student number, and room number~~. If you cannot change your name before you enter the Zoom meeting, you can change your name as follows once you entered the meeting:
  - Go to "Participants"
  - Find yourself
  - Click on "Rename"
  - Enter a screen name
- You will be prompted that the host wants to view your desktop. Choose the option "Start Sharing" and **share your Entire Screen**, i.e. we should see anything you open, and not just specific software or webpage tabs.
- You should see a green box around your screen indicating that you are sharing the image.
- Keep sharing your screen during the full duration of the exam.
- You do not need to show yourself via the webcam, we can see you via in-person invigilating.

Assignment instructions

Download materials (.zip):

- Download the folder "DataAndTemplate_GradedAssignment.zip" for this assignment from Brightspace and save and unzip it in a separate folder on your computer.
- Rename the folder to your student number.
- **Save the template with your student number as the file name, i.e. "studentnumber.Rmd"**.
- Include your student number as author name at the top of the template file.

R code Reproducibility:

- Write your code in the provided templates and make sure that it is reproducible code, e.g.
  - load all required packages,
  - do not rename the data files,
  - do not use folder paths that are specific to your laptop. If a new folder is created within the template folder, you may use paths that are relative to the folder containing the .Rmd file.
- Write your code in a structured way, e.g.
  - Packages should be loaded at the top of the script, so it easy to see which ones are needed to run your code,
  - Make sure that your code is easy to read for others, e.g. use spacing, give your variables concise but informative names, and include comments if needed.
- If you download more data, make sure the data is in the Data subfolder and specify that folder when you load the data.
- Do not remove any of the files in the folder and do not move any files.
- **Regularly save and knit your .Rmd file to check whether it produces a .pdf/.html file.**.

Submit your work (upload .zip file):

After the end time of the exam, you have 10 minutes to upload your files. If you miss this deadline, your work will not be graded. So, do not use these 10 minutes to continue working on the assignment, but use it to submit your files:

- Save and knit your .Rmd file to produce your final .pdf/.html file for submission. This file should be saved in the same folder as your template.
- Compress your folder (with your student number as its name) to a .zip-folder.
- Upload the .zip-folder to the Assignment submission portal on Brightspace.

Reproducibility (5%)

Follow all submission and code instructions, and make sure that your code is reproducible, i.e., we can knit your Rmd and produce the same .html or .pdf file (see tips in instructions).

Task 1. Data explorations (45%)

Make the plots as described in each of the subquestions of this task. Each of the plots should be self-contained (i.e. readers should be able to understand them without extra explanation) and obey the principles of good graphics.

Task 1a)

Emission class of vans, 2020 - 2024

Use the dataset on the emission class of vans in the Netherlands in 2020, 2022 and 2024 (`VanEmissionsLong.csv`). It contains the following variables:

- `Year`: Year of measurement (January 1st)
- `EmissionClass`: Emission class; indicates the level of air pollution of a van; the higher the number, the cleaner the vehicle.
- `Percentage`: Percentage of vans in this emission class

Create a suitable data visualization that shows how the distribution of the emission classes of vans changed from 2020 to 2024. The plot should focus on the observation that the percentage of vans of a higher (i.e. better) emission class increased in these years.

Task 1b)

Speed tickets in each province over 2023

Use the dataset on the number of speed tickets that are issued in the months of 2023 in each of the provinces in the Netherlands (`speedTickets_perMonthPerProvince.csv`). It contains the following variables:

- `Province`: Province in the Netherlands (12 options)
- `Month`: Month in 2023
- `Count`: Number of speed tickets issued
- `geometry`: Simple features data for each of the provinces in the Netherlands

Create a suitable data visualization that shows how the number of speed tickets changes over time for each province. From the plot, it should be clear which of the provinces generally have the highest number of issued speed tickets.

Task 1c)

Marvel film database

The `Marvel` data set ('Marvel.csv') contains data on several Marvel films, like the critics and audience scores and financial information (budget, gross, etc.). The variables that you need for this assignment are listed in the table below.

- `Film`: film title
- `Category`: category
- `WorldGross`: worldwide gross income (million dollars)
- `CriticsScore`: critics % score
- `AudienceScore`: audience % score

Create a suitable data visualization that shows whether the audience scores the marvel movies influence their worldwide gross income. Add the critics score and movie category as additional variables to the plot to give an indication of whether the opinion of the audience is in line with that of the critics, and whether the category of the movie matters.

Task 2. Remake the plot (50%)

Solar energy potential of the USA

The data visualization website *Information is Beautiful* created a series of visualizations exploring the United States' potential to reach net-zero carbon emissions by 2050. One of the plots is a cartogram heatmap showing the potential utility of solar power in each of the states.

Your task is to remake the data visualisation below, which is a choropleth map version of the cartogram. Hence, in this task you will plot the same data, but as a choropleth map. Note that the map only focuses on contiguous United States, i.e. 48 states and the District of Columbia, and excludes the states Alaska and Hawaii.

Use the dataset `potentialsolarenergyUSA.shp` in the data subfolder `potentialsolarenergyUSA_shp`.

The variables in the dataset are:

- `State`: State name
- `SlrPtnt`: Potential utility-scale capacity for solar power (terawatt hours)
- `geometry`: Simple features data for each of the states of the United States of America