

Digital Methods: Learning Journal

Jonas Nebel

Autumn 2019

1 Today's Date

1.1 Thoughts / Intentions

1.2 Action

1.3 Results

1.4 Final Thoughts

2 31/10/2019

2.1 Thoughts / Intentions

09:30am: Currently getting introduced to this whole overleaf thing. As a non-technical guy this all seems pretty overwhelming. I do hope everything clears up in due time.

10:00am: I think I got my GitHub working. Seems like an amazing program once you learn how to use it. I successfully created my own repository.

10:25am: Got both my GitHub and Overleaf working, which is a victory in itself.

10:50pm: After having complications with Overleaf, we are now looking at spreadsheets and how to use them to collect and sort data

2.2 Action

- Went to Overleaf to prepare the learning journal for future use
- Entered into Introduction for Data Organization in Spreadsheets
- Created a repository on github

Creating a repository on Github:

- Click on profile icon in top right corner and click 'your repositories'
- Click 'new' and entering details - named repository 'learning-journal'
- Selected initialize this repository with a README and created repository

Overleaf:

- Used sample template in Overleaf and edited title/abstract - recompiled and downloaded pdf for Github.

Data Carpentry Exercises:

Messy Spreadsheet

- Opened data sets from previously downloaded excel spreadsheets
- Found 'tabs' in excel in bottom left corner
- Discussing messy data
- Color used for indication
- Multiple variables in column and row
- Multiple values in single cells
- Inconsistencies in null, 0 and false in each tab

- Inconsistencies and use of asterisk
- Blank cells are not adequate indicators
- Multiple tabs in one spreadsheet

Metadata

- Unclear data
- There seems to be a reoccurring and limited representation of items owned. Is there a specific criteria for items to be recorded?
- What does no to member associations exactly mean? That the members in the household are recluse or separate from the society they are living in?
- What is the question and context for 'affect-conflict' - what affects the conflict?
- Does years-liv refer to years lived in the house or community?

Formatting Problems

- Use 'blank' for a null value - best option
- Use *(underscores) for spaces in between words i.e. formatting_problems*

2.3 Results

- Github - new repository named learning-journal with practice upload committed.
- Overleaf - very confused, do not have a document to work with that is in pdf format. Leaving for now, will continue journal in cloudstor and will ask for assistance in class.
- Data Carpentry - straight forward and informative exercises. Completed up until Formatting problems.

2.4 Final Thoughts

10:55am: Using github and overleaf will take a lot more practice and hopefully some assistance.

3 07/11/2019 Hands-on session

3.1 Thoughts/Intentions

- **08:00am:** Getting introduced to Data cleaning with Open Refine
- **09:00am:** Quizzing in regex
- **09:15am:** Now starting to work with OpenRefine. I honestly feels terrifying.

3.2 Action

Scoping Exercise

- Opening OpenRefine
- Navigate to datacarpentry.org
- Download files for the lesson
- Getting data into OpenRefine
- Transforming data with []
- Filter data by using indicators such as `!`
- Learning to save work as script
- Import and compare scripts
- Completed document and saved as PDF
- Exporting our script so others can use them and replicate our work

3.3 Final Thoughts

10:50am

- Coding seems like a smart way to compile and understanding huge amounts of data. OpenRefine as a program showed me that big data sets is not necessarily out of scope.

4 14/11/2019

4.1 Thoughts/Intentions

08:15am: Getting introduced to Shell and Git(Bash).

8:40am: Terminal, as the name is on Mac, is a really intimidating program. It feels like I am violating the very core of my PC.

4.2 Action

- Introduction to Shell
- Navigating in different directories
- The command: `pwd` lets you see your current directory
- Creating copying and deleting files
- Redirecting commands to a file with pipes
- Introduced to Git.
- Git is a tool to manage your source code history.

4.3 Final Thoughts

10:55am The further we got into Shell, the less intimidating it felt. It felt extremely powerful, and useful in some situations. Still scary that you can erase your entire PC from there. Git is a strong tool to track changes in your computers files.

5 21/11/2019

5.1 Thoughts/Intentions

08:15am: Have downloaded R and RStudio in preparation for this class.

8:30am: Really excited to learn R, seems like it have a lot of possibilities.

5.2 Action

- Setting up R and learning to navigating it.
- Installing different packages.
- Learning about different arguments and assignment operators like `=`
- Assigning new values with vectors
- Importing and getting information out of CSV files
- Convert factors and renaming them

5.3 Results

- Setting up R and learning to navigating it.
- Able to get data into R, and manipulate it. Can make my own simple scripts.

5.4 Final Thoughts

10:55am The further we got into Shell, the less intimidating it felt. It felt extremely powerful, and useful in some situations. Still scary that you can erase your entire PC from there. Git is a strong tool to track changes in your computers files.

6 28/11/2019

6.1 Thoughts/Intentions

08:15am: Exited to go deeper into R territory to see what I have to offer

8:30am: Especially looking forward to the graphical visualisations.

6.2 Action

- Setting up R and learning to navigating it.
- Data manipulation with dplyr and tidyr
- Both accessible with `/library(tidyverse)` package
- Filtering rows and columns for specific information
- Selecting rows/columns and filtering at the same time with pipes `%>%`
- Mutating variables to find ratio of values in two columns
- Using ggplot to plot the relationship of two variables into a data visualisation
- Learning to do all different kind of plots
- Multiple variables in the same plot

6.3 Results

- I am now able to manipulate rows and columns
- I am able to do a visual presentation of my data
- Had a hard time getting the multiple variables in one plot to work

6.4 Final Thoughts

10:55am The deeper we get into R, the more comfortable I get with scripting. At the same time, I also feel like there is so much more to learn. I definitely have to practice the multiple variable integration in one plot.

7 28/11/2019

7.1 Thoughts

06:15am: Currently looking around the internet for inspiration to what I should write my final paper on. I stumbled across this article: <https://au.dk/univers/nyhed/artikel/fedme-er-ogsaa-et-socialt-problem/> which claims that obesity is not just a biomedical problem, but also a social one. The society, according to Anders Lindelof, promotes obesity. Having struggled with overweight most of my life, this struck something. I remembered reading this article in high-school: <https://faktalink.dk/titelliste/overvaegtige-boern-og-unge> that states that obesity amongst children has more than tripled over the last 30 years. It was at this point I decided I wanted to do something on the historical development in child obesity.

8 01/12/2019

8.1 Thoughts

07:15am: Still wondering what direction to take my paper in. Going to www.ourworldindata.org too see if I can get some inspiration.

07:45am: Found research and data on childhood obesity which I can use for a multi-country comparison.

09:20am: Now that both the society and the the biomedicin has been pointed at as sinners for childhood obesity, I wondered whether there could be an economic factor, seeing as childhood obesity still seems to be a problem, which would block for the efforts made to improve other sectors in preventing this issue. More specifically, I would like to investigate whether a country's wealth over time have an effect on child obesity.

15:15am: Found data on www.ourworldindata.org that shows the development in GDP per capita over time, which I have decided I want to use for my research.

9 08/12/2019

9.1 Thoughts/intentions

06:15am: Finally got around to start working on my script

06:25am: Downloaded both the CSV files containing data on childhood obesity and GDP per capita.

06:30am: Trying to get as much work done today as possible, since my son is hospitalized.

9.2 Action

- Created a project in R

- Made two scripts in R
- read a different CSV file in each script
- The following procedures was replicated in both scrips
- Started by assigning the data to a new value/variable that is easier to work with
- Extracted data from the six countries I chose to analyze and assigned a a new value to each country, by filtering them
- Inspected them to see that they all had equal amounts of comparable data
- I then wanted to collect the data from the 6 countries in one variable. I encountered my first problem here
- ERROR: I started by seperating the countries I wanted to use with commas.
- SOLUTION: I went back to the carpentry exercises and learn I had to use — to separete them use the column name every time a new country was to be added.
- The GDP data did not go from 1990-2016 per default like the data with obese children, so here I had to filter out the years I wanted to use. First i tried to show years from 1990-2016 only wiht the =j command. This worked, however the dataset still had observations for every year, and I needed observations from every 5th year.
- This was done by doing that same command that collected the countries in one value, using the Year Column instead, and what years I wanted extracted.

9.3 Results

- A new dataset containing information about childhood obesity for the six countries in question
- A new dataset containing information about GDP per capita for the 6 countries in question

9.4 Final Thoughts

21:30am Coding and making a script take a lot longer than initially realized. It does however really ease the process of large quantities of data.

10 26/12/2019

10.1 Thoughts/intentions

07:00am: Want to finalize the script

07:05am: Make a graphical illustration with ggplot containing the timeline of all 6 countries

06:30am: Do a correlation test

10.2 Action

- read up on the datacarpentry about ggplot
- Trying to portray my ggplot as a timeseries with lines
- ERROR: Would not let me make a ggplot using the line command.
- SOLUTION: After a quick conversation with Antonio Rivero he made me realise that I needed to change the absurdly long Column name, since it was filled with expressions that means something different in R-language.
- After changing my Column names to something more sensible, I tried making the ggplot again.
- ERROR: Still would not work as I wanted it to.
- SOLUTION: I asked my good friend Kristian to take a look at it, since he has experience in another code language, Python. He told me to try the "smooth" function and adding color to my country colum. This worked like a charm!
- This was done in both scripts.
- Now, having created a graphical data visualisation, I wanted to see whether there was any correlation between the two or not.
- I did now now how to make a correlation or even if it was possible between scripts.
- I found this youtube video: [youtube.com/watch?v=xsL4yLBNyDg](https://www.youtube.com/watch?v=xsL4yLBNyDg) which explained how to do it.
- The test showed an overwhelming correlation, which, after all, is very hard to generalize due to the small sample size.

10.3 Results

- A graphical data visualisation showing the development in a timeline from 1990-2016 for all 6 countries concerning childhood obesity.
- A graphical data visualisation showing the development in a timeline from 1990-2016 for all 6 countries concerning GDP per capita.
- A quite clear correlation between the two variables, which suggest that GDP per capita have quite the impact on childhood obesity. Even though this is hard to generalize, it does at least bring awareness to this correlation, in deciding whether or not more research should be done on this.

10.4 Final Thoughts

23:30am Being finished with the scripts is really satisfying. It has really opened my eyes up for how much you can achieve by having a little finesse when it comes to coding. This has definitely been an eyeopening experience for me. I am pretty sure I will expand on this knowledge in the future.