NEPTTP WITS Data Carpentry Workshop

29-31 January 2020
Instructors: Angelique van Rensburg, Juan Steyn

Collaborative document: https://pad.carpentries.org/WITS-DC

Workshop website: https://sadilar.github.io/2020-01-29-WITS/

post workshop - https://carpentries.typeform.com/to/UgVdRQ?slug=2020-01-29-WITS

User login: vmuser

Breaks 10 / 10:30

Takunda/BA Honours in Development Studies/786359@students.wits.ac.za

Fritz Nganje/DLitt et Phil Politics/386496@student.wits.ac.za

Babalwa/ BA Honours in Political Studies/848935@gmail.com

Ayton Altorio / MA Demography & Population Studies / 1148071@students.wits.ac.za

Khanyisile Ntsenge/PDM Monitoring and Evaluation/480274@students.wits.ac.za

Leslie Dwolatzky/MA African Politics/715348@students.wits.ac.za

Lucas Banda /MA DEMOGRAPHY AND POPULATION STUDIES/764745@students.wits.ac.za

Kholonyane Quincy/ BA Honours in International Relations/ 1381077@students.wits.ac.za

Survey: https://carpentries.typeform.com/to/wi32rS?slug=2020-01-29-WITS

Things to download:
1. Data for first lesson: https://librarycarpentry.org/lc-spreadsheets/data/training_attendance.xlsx

Group 1

Group 2
PGR/PDRA/Other Column does not make it easy to separate the information for analysis.
Length in the open access is written both in numbers and words

Group 3
Hours in second table recorded as text
3 elements recorded in third column will be treated as 1

Doesn't start in Row 1 column A
Acronyms don't make sense
Would be easier to convert text data to numerical meanings in "delivered by" column

Spreadsheets:
1. How to do data validation in Libre Office
https://libreofficehelp.com/data-validation-using-dropdown-list-libreoffice-calc/
https://www.youtube.com/watch?v=uMin77n_flE
https://www.youtube.com/watch?v=Jstu2a4F0yI

More resources for data validation:

https://datacarpentry.org/spreadsheet-ecology-lesson/04-quality-control/index.html

Open Refine
https://github.com/OpenRefine/OpenRefine/releases/download/3.3-rc1/openrefine-linux-3.3-rc1.tar.gz

https://github.com/LibraryCarpentry/lc-open-refine/raw/gh-pages/data/doaj-article-sample.csv

If you closed your browser just type this into your browser
127.0.0.1:3333

"https://api.crossref.org/journals/"+value

R

- What you liked

I wanted a challenge in a new degree and I sure got it. But I definitely appreciated the help that was given to us

hands on approach to instructions

Very Impressed with how excel works. I only need to remember every principle taught today, Thanks.

the content was explained well and the pace was good.

Interactive and fun. Very interesting.

Open Refine looks very useful and it was a great introduction to it

Open refine is a great introduction to the new degree programme

- What could be improved

I definitely want to study at home further so that I can be confident with my work

perhaps a little bit slower

regular, smaller breaks

There could have been less of a rush. But I think its up to us to do our own reading and practice at home to try and catch up on what we might have missed or not understood.

Not too sure how to recap on these exercises at home because the teaching was a little fast

More exercises

The practical side of all these lessons are too fast and once you got lost, you are lost forever.

What are the values after each statement in the following?
```
mass <- 47.5          # mass?
age  <- 122           # age?
mass <- mass * 2.0    # mass?
age  <- age - 20      # age?
mass_index <- mass/age  # mass_index?
```

download.file(url="https://ndownloader.figshare.com/files/2292169",destfile = "data_raw/portal_data_joined.csv")