# Sixth exercise class

Class 5

Introduction to numerical programming and analysis

Asker Nygaard Christensen

Spring 2021

## Plan

1. Next assignment

2. Reviews and poll

3. Notes on problem set 3

# Next assignment

## Data Analysis Project

It might already be a good idea, to start thinking about what you wanna do and what kinda data you wanna use in the next assignment You can see the assignment at Github, but basically you have to download data from an online source and do some empirical analysis (key figures and graphs). You get to choose the data and what kind of analysis you wanna do yourself. There are LOADS of possibilities, so choose something you find interesting and seems manageable. Recreating a already existing figure using python is also good. You can see all the previous projects by searching 'projects-2019' and 'projects-2020' at the NumEconCopenhagen github account. Also if you wanna do some regressions you can use statsmodels, but regressions are not a prerequisite for doing the assignment, a beautiful figure is just as good. Since these projects are idiosyncratic, you'll find less directly applicable code in the lectures and PS. But they are still good for inspiration, along with googling your problems.

## Ideas

- Investegating the home-field advantage using Covid-variation. <u>This</u> should be able to the job. But you can also download excel sheets from <u>here</u> (Inspiration from <u>this paper</u>)
- Twitter data (See for example <u>This guide</u>
- Google trends using <u>pytrend</u>
- You recreate graphs from 'Capital in the 21st century' and 'Captial & Ideology', either from <u>QUANDLE</u> or <u>Piketty's personal website</u>
- <u>World Inequality Database</u>
- <u>Pandas_datareader</u>, from L08 has the capability to load from multiple sources, including the World Bank, OECD and stock data.
- Denmarks statistics can be accessed using pydst, also L08
- There is even a python package for downloading <u>IMDB data</u> (A group actually used IMDB data last year, <u>Credible threats</u>, although they downloaded the data manually)

## My tips on Pandas and data science generally

Data science can be an excruciating job. Because you're doing self-chosen projects, you won't be able to rely as much on the lecture and PS. Remember to reach out when you're having problems.

There are loads of python guides and stack-flow answers on the internet, so the right google search can also be your saviour.

You been introduced to many different ways of referencing data. I'd recommend using *.loc[I,columns]* mostly in the beginning, as it is the most versatile. Instead of the conditions implicit in *I*, you can also use a list of index-numbers.

When creating Boolean condition, I'd also recommend being explicit with your brackets. Also, remember '*&*' is the bit-wise 'and' operator, and '|' is the bit-wise 'or' operator (Depending on your keyboard is could be AltGr+'the key with | on it' (close to backspace), or Alt+i), also called pipe or vertical line.

# Reviews and poll

# Notes on problem set 3

## My notes on problem set 3

I'd would recommend to be more lenient about looking at answers in this problem set, as it mostly focuses on referencing and adjusting data, and is therefore a bit syntax heavy. You still need to make sure you understand what the syntax is doing though!.

In 2.5, when using the .drop() method, remember to use the argument *inplace=True*, to alter the DataFrame you're calling it on.

In 2.6 it says rename 'consumption' to 'cons', but it should just be 'con'.

At problem 2.7, beyond the lines you are explicitly asked to correct, there is an additional error in the *try:*-block, at the '*dt['assets_3']* =...' - line.

Also 2.7 is quite hard, because you have to get all 4 correct at the same time, and you don't get a error message. A suggestion would be to run the lines individually to check if they are correct.

## My notes on problem 3

For loading excel data in problem 3, look at point 3.1 in lecture 7.
And remember when doing this exercise to continuously look at the data, to see if the operations you have performed worked as expected.
And what you need to do to make it look like answer.
Notice that the *rename_dict* does not contain 'Unnamed: 0', which contains year, you need to add it yourself.