Project Proposal № 1
COSC3000: Visualization, Computer Graphics & Data Analysis

William E. G. Kvaale, University of Queensland   13/03/2020

What

For my visualization project I will be investigating the COVID-19 pandemic currently disrupting the world.

My data source will be obtained from the daily updated datasets provided by Johns Hopkins University Center for Systems Science and Engineering (CSSE).

The goal for my project is to concisely present the current situation. I will also look into obtaining other reliable data sources for previous viruses such as SARS, MERS and/or swine flu. Hopefully my analysis can put the current situation in to a comparable context.

Why

The situation is affecting everyone's lives, including my own. This motivates me to look deeper into what is actually going on. And by attempting to create good visualizations representing the situation and comparing it to past virus outbreaks, I might gain some new insight and test my knowledge and skills in regards to data visualization.

How

Mainly I will be utilizing Python with Pandas. The data comes in the form of Comma-Separated Values files, and I will therefore use Pandas' read_csv to structure it into DataFrames. From this point on the data is more manageable for analysis.

Further I will most likely use NumPy for any computation in regards to matrices or vectors. For visualization I will rely on Matplotlib and/or Seaborn to do the heavy lifting. For any geospatial data I will work with GeoPandas.

Listing 1: Loading COVID-19 Dataset with Python

```python
1  # Import Modules
2  import pandas as pd
3  from os import path
4
5  # Load Dataset
6  def load_data(filePath):
7      df = pd.read_csv(filePath)
8      # perform whatever preprocessing of the data needed
9      return df
10
11 # Declare path to dataset
12 WHO_COVID_19_SIT_REP_TS = "dataset/csse_covid_19_data/←
       csse_covid_19_daily_reports/03-11-2020.csv"
13
14 data = load_data(WHO_COVID_19_SIT_REP_TS)
15 # Show 10 first rows
16 data.head(10)
```

Figure 1: Sample data from the CSSE COVID-19 Daily Reports, 11.03.2020

	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered	Latitude	Longitude
0	Hubei	China	2020-03-11T10:53:02	67773	3046	49134	30.9756	112.2707
1	NaN	Italy	2020-03-11T21:33:02	12462	827	1045	43.0000	12.0000
2	NaN	Iran	2020-03-11T18:52:03	9000	354	2959	32.0000	53.0000
3	NaN	Korea, South	2020-03-11T21:13:18	7755	60	288	36.0000	128.0000
4	France	France	2020-03-11T22:53:03	2281	48	12	46.2276	2.2137
5	NaN	Spain	2020-03-11T20:53:02	2277	54	183	40.0000	-4.0000
6	NaN	Germany	2020-03-11T19:13:17	1908	3	25	51.0000	9.0000
7	Guangdong	China	2020-03-11T10:13:06	1356	8	1282	23.3417	113.4244
8	Henan	China	2020-03-11T08:13:09	1273	22	1249	33.8820	113.6140
9	Zhejiang	China	2020-03-11T09:33:12	1215	1	1195	29.1832	120.0934
10	Hunan	China	2020-03-11T02:18:14	1018	4	995	27.6104	111.7088

Data sources

I will download my data from the daily updated GitHub repository from Johns Hopkins CSSE. The repository is open to the public, and can be found here. The CSV-files can also be obtained by using cURL in the following way:

Listing 2: Loading data using cURL.

```
1 curl -LJO https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master↩
      /who_covid_19_situation_reports/who_covid_19_sit_rep_time_series/↩
      who_covid_19_sit_rep_time_series.csv
```