Exploring the relationship of multiple factors on the spread of COVID-19 in countries

FIT5147 - DATA EXPLORATION & VISUALIZATION – S1 2020

POURIA EBRAHIMNEZHAD

The Questions

1. Is there any relationship between the spread of COVID-19 and the median age of the countries impacted?

2. Is there any relationship between the spread of COVID-19 and the population density of the countries impacted?

3. Is there any relationship between the spread of COVID-19 and different health and governmental factors (i.e. smoking, life expectancy, Government spending on health)?

Data Sourcing & challenges

Data Sources:

- COVID-19 Data (OCHA - HUMANITARIAN DATA EXCHANGE)
- Age, life expectancy, Government health expenditure, smoking, population density (The World Bank – worldbank.org)

Wrangling Challenges:

- Missing data
- Making one csv fit for purpose (country names, reshaping data, etc.) – Python, Excel

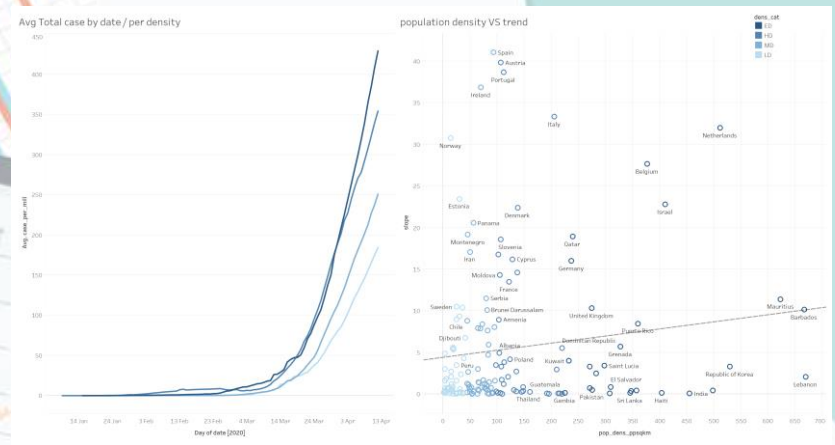
Data Exploration challenges:

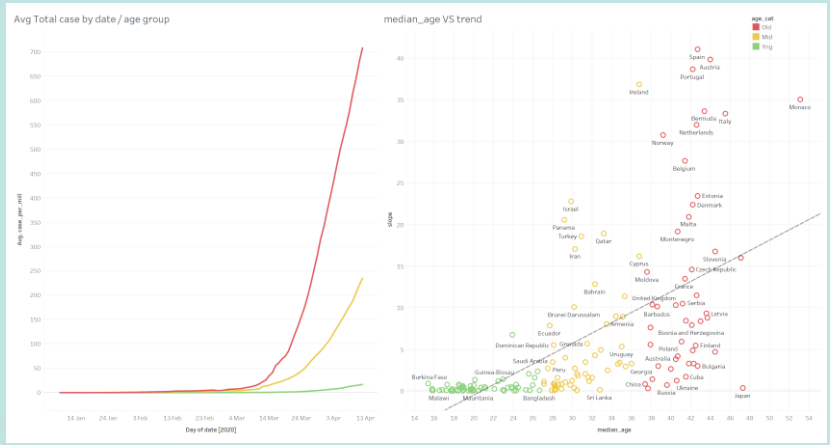
- Defining a suitable metric to measure spread of disease in countries – (slope of best fit linear model)
- Levelling the comparison field – (case per million)
- Eliminating outliers to allow for comparison – z score statistical test (Andorra, Iceland, Switzerland, etc.)

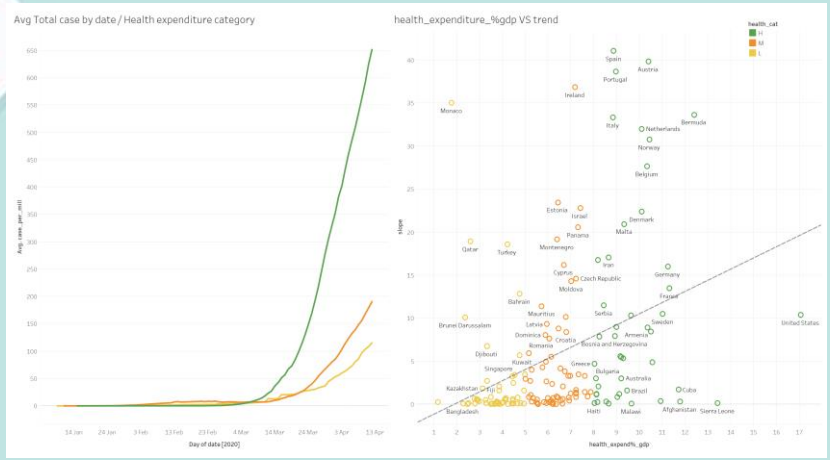
Final Format

new_id	country	date	cum_conf	new_case	cen_long	cen_lat	median_ag	pop_dens_	life_exp	smoking_%_adults	health_exp	latest_tota	case_per_r	slope	age_cat	dens_cat	life_cat	smoke_cat	health_cat
1	Afghanista	25/02/2020	1	1	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
2	Afghanista	26/02/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
3	Afghanista	27/02/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
4	Afghanista	28/02/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
5	Afghanista	29/02/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
6	Afghanista	1/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
7	Afghanista	2/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
8	Afghanista	3/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
9	Afghanista	4/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
10	Afghanista	5/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
11	Afghanista	6/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
12	Afghanista	7/03/2020	1	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.026902	0.265567	Yng	MD	L	H	H
13	Afghanista	8/03/2020	4	3	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.107607	0.265567	Yng	MD	L	H	H
14	Afghanista	9/03/2020	4	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.107607	0.265567	Yng	MD	L	H	H
15	Afghanista	10/03/2020	4	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.107607	0.265567	Yng	MD	L	H	H
16	Afghanista	11/03/2020	4	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.107607	0.265567	Yng	MD	L	H	H
17	Afghanista	12/03/2020	7	3	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.188312	0.265567	Yng	MD	L	H	H
18	Afghanista	13/03/2020	7	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.188312	0.265567	Yng	MD	L	H	H
19	Afghanista	14/03/2020	7	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.188312	0.265567	Yng	MD	L	H	H
20	Afghanista	15/03/2020	10	3	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.269017	0.265567	Yng	MD	L	H	H
21	Afghanista	16/03/2020	16	6	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.430427	0.265567	Yng	MD	L	H	H
22	Afghanista	17/03/2020	21	5	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.564935	0.265567	Yng	MD	L	H	H
23	Afghanista	18/03/2020	22	1	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.591837	0.265567	Yng	MD	L	H	H
24	Afghanista	19/03/2020	22	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.591837	0.265567	Yng	MD	L	H	H
25	Afghanista	20/03/2020	22	0	66.02653	33.8389	18.8	56.93776	64.13	35.2	11.77719	37172386	0.591837	0.265567	Yng	MD	L	H	H

Data Exploration

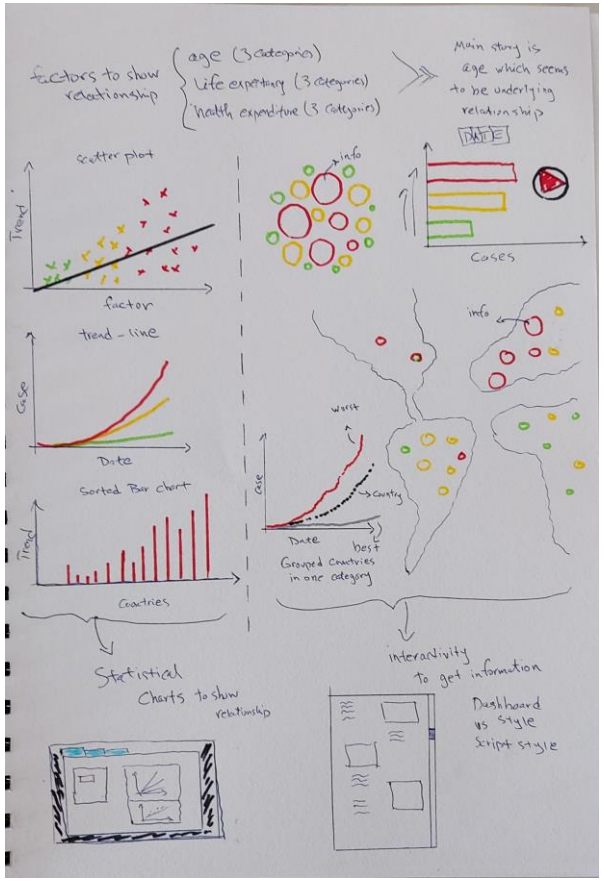






Brain Storm Ideas

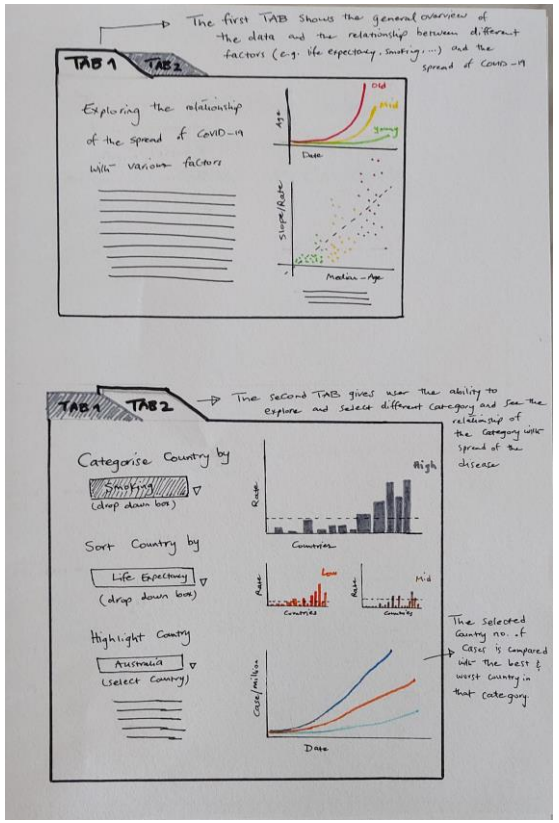
- Demonstrate relationships
- How much Interactivity
- Categorise visualisations
- Wild ideas



Initial Design 1

- Advantages:
  - Complete EDA info
  - Space for narrations
  - Multiple graphs

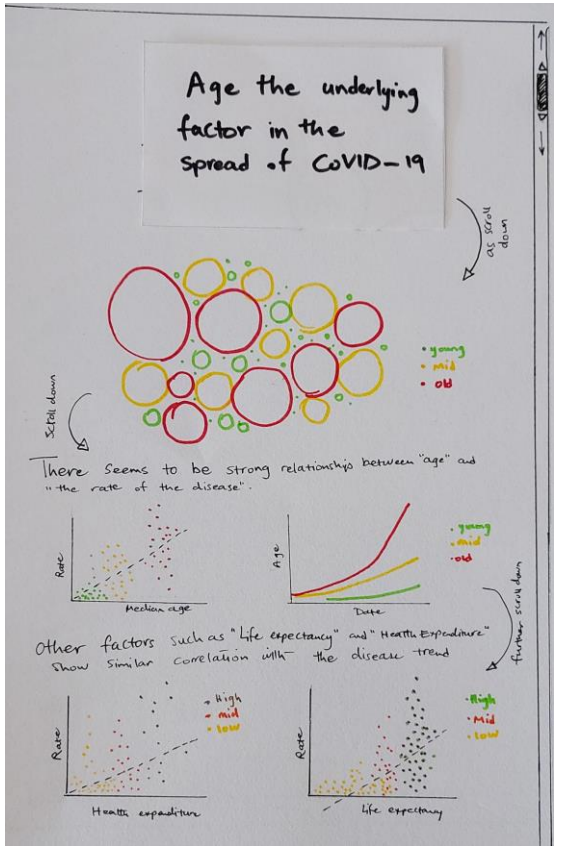
- Disadvantages:
  - Too interactive
  - Requires user to explore
  - Redundant to story line



Initial Design 2

- Advantages:
  - Good story telling
  - Take the user step by step

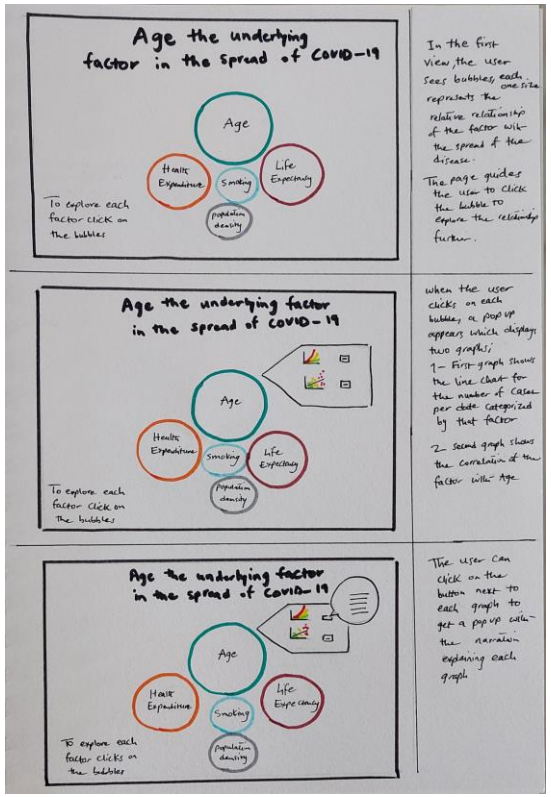
- Disadvantages:
  - Not much interactivity
  - Difficult to implement
  - Less innovation



Initial Design 3

- Advantages:
  - Good holistic view
  - not too complex to view

- Disadvantages:
  - Not much room for text narration
  - Difficult to implement



Final Design