Project Title

World Health Organization Mortality Investigation

https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

Team Members

Bryan Arment
CSCI
University of Colorado Boulder
Boulder CO USA
brar9262@colorado.edu

Jeff Dank
CSCI
University of Colorado Boulder
Boulder CO USA
jeda7205@colorado.edu

Tony Pearo
CSCI
University of Colorado Boulder
Boulder CO USA
anthony.pearo@colorado.edu

Questions Sought To Answer

- How does mortality present itself in countries across the world?
- Can we make any predictions about future mortality?
- Can we identify groups of countries with similar mortality profiles?

Data Preparation Work

1) Discover a dataset with over 1 million entries

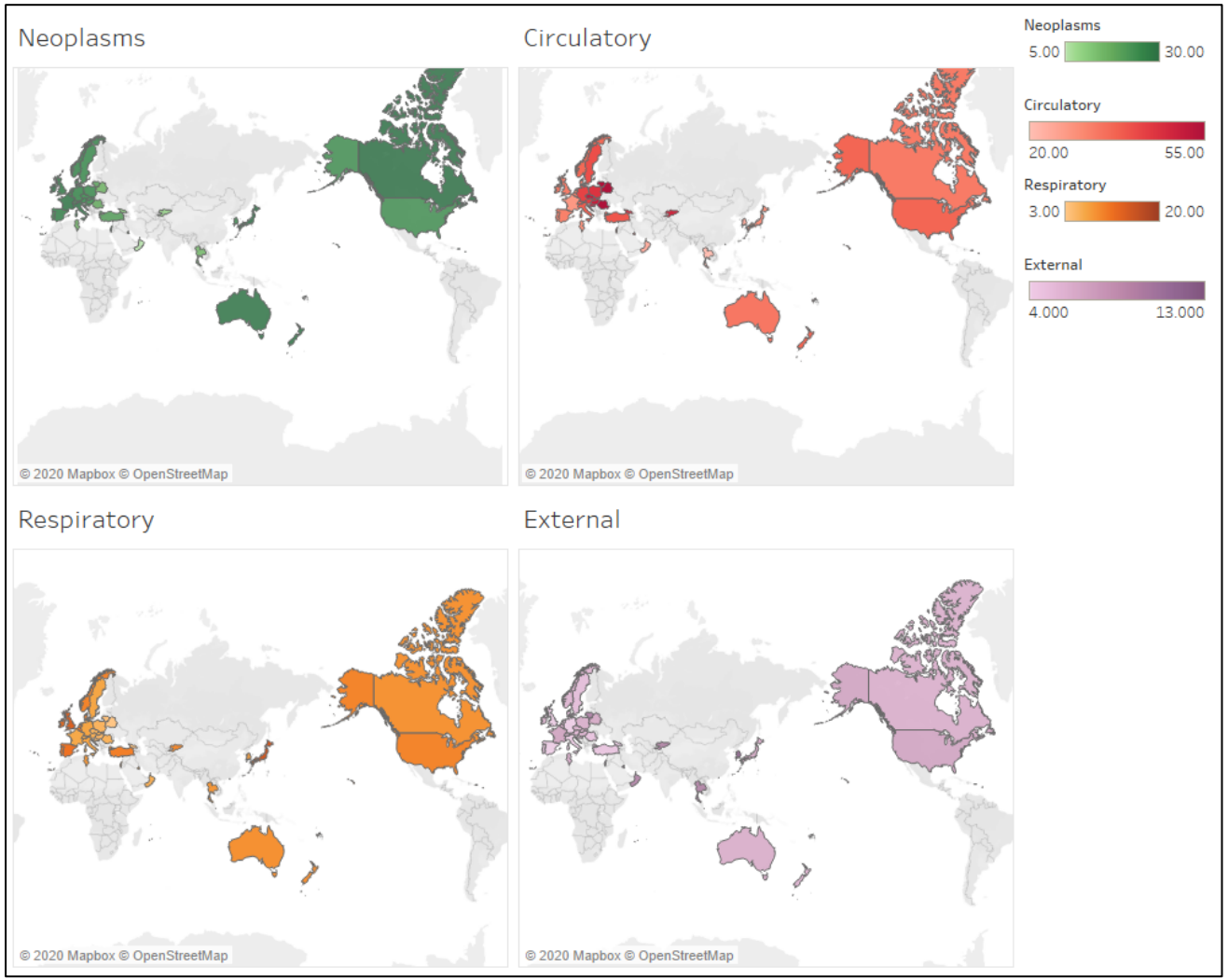
2) Merge the data into one data frame, drop bad values, normalize deaths by population

3) Re-group the over 9000 causes of deaths into 22 distinct groups

4) K-means, EDA, plotting

Tools Used

- Python / Jupiter notebooks
  - Pandas library
- Excel
- Tableau



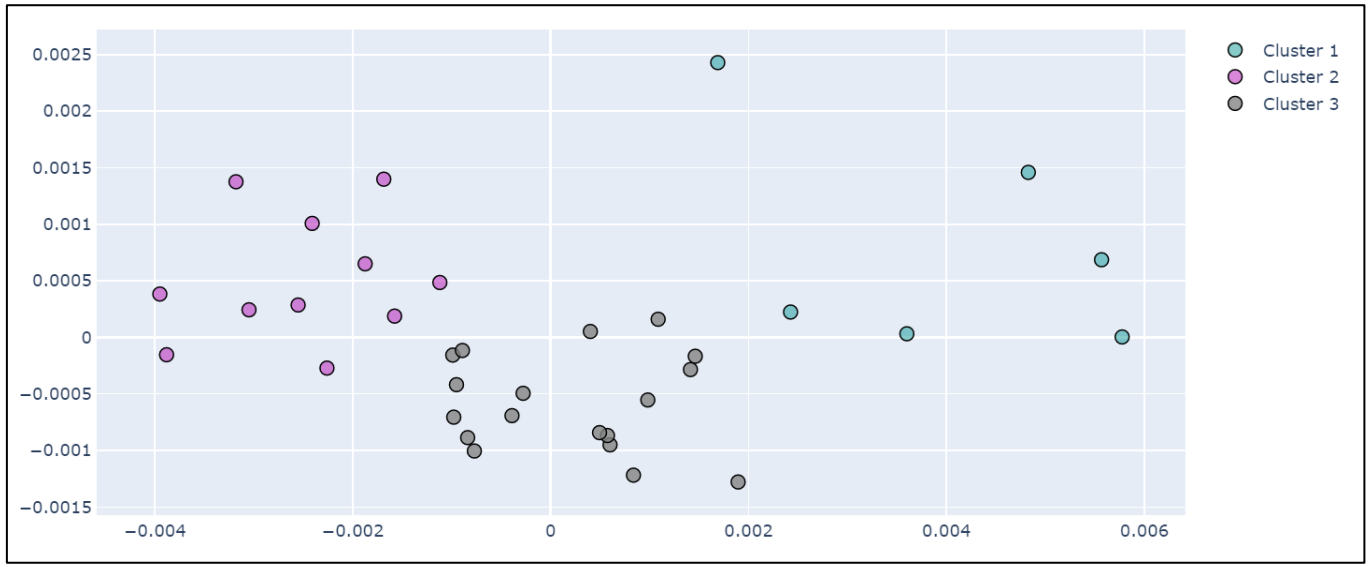
Classification/clustering/etc applied

Top 10 Averaged Normed Death Rates

	Category	Disease	Average Normed Death Rate (ANDR)	Cluster 1	Cluster 2	Cluster 3
21	22	All Together	0.008897	0.011717	0.006728	0.009283
8	9	Circulatory	0.003665	0.006549	0.002536	0.003395
1	2	Neoplasms	0.002166	0.002396	0.001474	0.002512
9	10	Respiratory	0.000762	0.000543	0.000621	0.000921
19	20	External	0.000513	0.000770	0.000421	0.000484
10	11	Digestive	0.000386	0.000561	0.000255	0.000408
3	4	Endocrine	0.000310	0.000187	0.000445	0.000269
17	18	Unclassified	0.000232	0.000198	0.000187	0.000272
5	6	Nervous	0.000232	0.000133	0.000203	0.000283
4	5	Mental	0.000189	0.000058	0.000115	0.000277

Classification/clustering/etc applied

2005 K-Means Clustering Analysis



Interpretation of 2005 Cluster Data					
Cluster 1		Cluster 2		Cluster 3	
Country	Disease	Country	Disease	Country	Disease
Croatia	All Together	Mauritius	All Together	USA	All Together
Czech Republic	Circulatory	Rodrigues	Circulatory	Japan	Circulatory
Georgia	Neoplasms	Canada	Neoplasms	Austria	Neoplasms
Hungary	External	Cyprus	Respiratory	Belgium	Respiratory
Lithuania	Digestive	Hong Kong SAR	Endocrine	Denmark	External
Romania	Respiratory	Israel	External	France	Digestive
	Unclassified	Iceland	Digestive	Germany	Nervous
	Endocrine	Kyrgyzstan	Nervous	Italy	Mental
	Nervous	Luxembourg	Unclassified	Netherlands	Unclassified
	Genitourinary	Malta	Genitourinary	Norway	Endocrine
		New Zealand		Poland	
				Spain	
				Sweden	
				Switzerland	
				United Kingdom	
				England and Wales	
				Northern Ireland	
				Scotland	

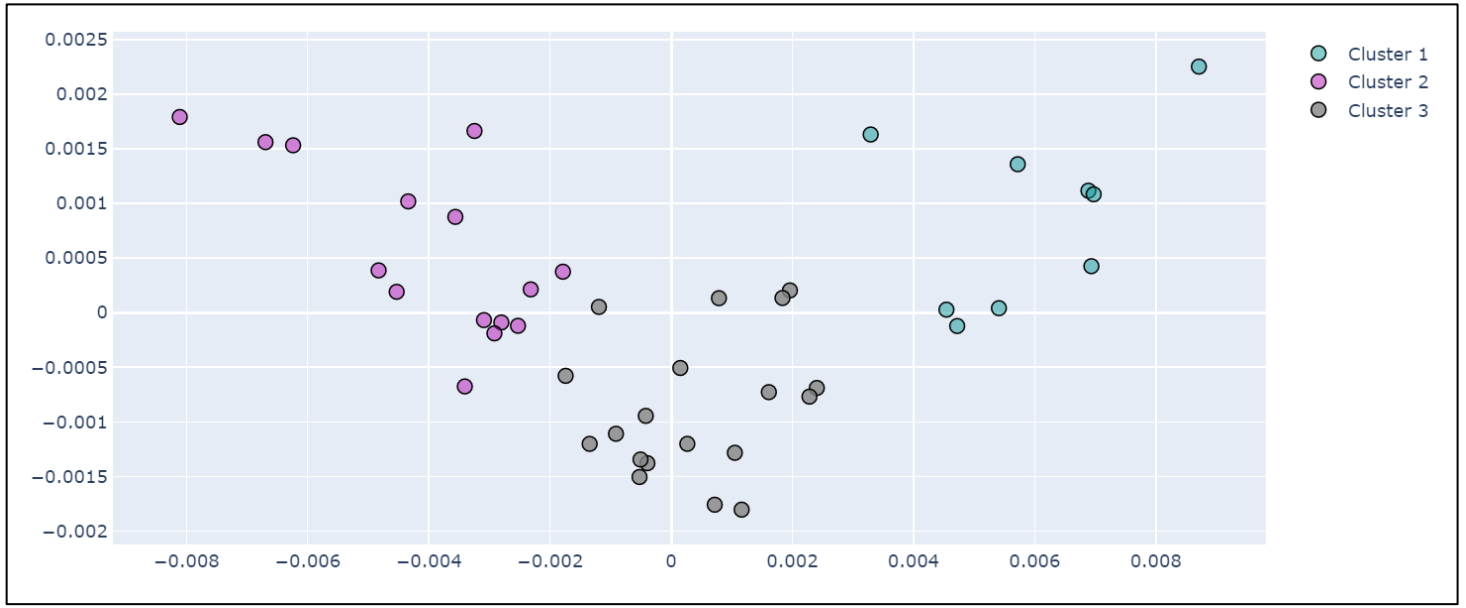
Classification/clustering/etc applied

Top 10 Averaged Normed Death Rates 2015

	Category	Disease	Average Normed Death Rate (ANDR)	Cluster 1	Cluster 2	Cluster 3
21	22	All Together	0.009057	0.013607	0.005527	0.009689
8	9	Circulatory	0.003598	0.007401	0.001838	0.003186
1	2	Neoplasms	0.002166	0.002707	0.001261	0.002625
9	10	Respiratory	0.000785	0.000604	0.000601	0.001017
19	20	External	0.000440	0.000649	0.000308	0.000445
10	11	Digestive	0.000370	0.000620	0.000204	0.000382
17	18	Unclassified	0.000339	0.000611	0.000136	0.000371
3	4	Endocrine	0.000319	0.000260	0.000394	0.000288
4	5	Mental	0.000284	0.000137	0.000149	0.000459
5	6	Nervous	0.000275	0.000196	0.000170	0.000395

Classification/clustering/etc applied

2015 K-Means Clustering Analysis



Interpretation of 2015 Cluster Data					
Cluster 1		Cluster 2		Cluster 3	
Country	Disease	Country	Disease	Country	Disease
Bulgaria	All Together	Mauritius	All Together	Japan	All Together
Croatia	Circulatory	Rodrigues	Circulatory	Austria	Circulatory
Georgia	Neoplasms	Cyprus	Neoplasms	Belgium	Neoplasms
Hungary	External	Hong Kong SAR	Respiratory	Czech Republic	Respiratory
Latvia	Digestive	Israel	Endocrine	Denmark	Mental
Lithuania	Unclassified	Jordan	External	Germany	External
Moldova	Respiratory	Kuwait	Digestive	Greece	Nervous
Romania	Endocrine	Maldives	Nervous	Italy	Digestive
Serbia	Nervous	Singapore	Mental	Malta	Unclassified
	Genitourinary	Turkey	Genitourinary	Netherlands	Endocrine
		Ireland		Norway	
		Kyrgyzstan		Poland	
		Luxembourg		Portugal	
		Australia		Spain	
		New Zealand		Sweden	
				United Kingdom	
				England and Wales	
				Northern Ireland	
				Scotland	

Knowledge Gained

- The top two groups of deaths are circulatory or neoplasms (growths / cancers)
- France has low circulatory deaths compared to its surrounding European countries
- Countries clustered in a pattern that aligned with our hypothesis.
- Defining a specific question can help break down a large dataset
- We trimmed a lot! Age, gender, NAN…

How Can That Knowledge Be Applied?

- Due to lower circulatory death rate investigate French diet, drinking water, air quality, other lifestyle habits
- General knowledge of risk factors
- Policy makers can look to more ideal healthcare systems.