Project 4 Group 2:

Government Healthcare Expenditure and Infectious Disease Mortality Rates

By: Kelly Carter, Ben Comfort, Barrett Fudge, and Cory Lingerfelt

Agenda

- Government Health Expenditure vs. Mortality Rates Specific
- Is there a Relationship?
- Predictive Machine Learning Algorithm
- How can we use Current Health Expenditures of Countries to Lower Infectious Disease Mortality Rates in the Future?

Assumptions

- The relationship between infectious disease mortality rates and government health expenditures as a percentage of their GDP will have a relationship with some level of linear correlation/potential causation.

- The higher a percentage a government spends on their country's healthcare, the lower their mortality rate will be on average.

Methodology

Initial Databases

Year	Number	Percent_of_Total	Age_Stnd_Per_100k	Death_rate_per_100k
2011	14572.0	4.600982	28.986910	35.265788
2012	13986.0	4.409984	27.442447	33.495233
2013	14654.0	4.532254	28.102552	34.728386
2014	14816.0	4.600742	27.769779	34.748745

1. WHO Mortality Database
   a. Infectious and Parasitic Diseases Only
   b. Cleaned and filtered data in python
2. WHO Global Health Expenditure Database
   a. 2010 to 2022
   b. Current Health Expenditure (CHE) and Capital Health Expenditure (HK)
3. Joined data
   a. SQL join and storage

	country	code	region	income	year	che_gdp(%)	hk_gdl(%)	che_pc_usd	che	gdp_pc_usd	che_usd
0	Argentina	ARG	AMR	Upper-middle	2011	9.4	0.0	1,203	205,225	12,769	49,931
1	Argentina	ARG	AMR	Upper-middle	2012	9.8	0.0	1,362	259,235	13,859	57,139
2	Argentina	ARG	AMR	Upper-middle	2013	9.8	0.0	1,415	327,497	14,469	59,988
3	Argentina	ARG	AMR	Upper-middle	2014	9.7	0.0	1,281	442,857	13,241	54,841

Exploratory Data Analysis: Predicting Mortality

Assumption:

- Linear Relationship
- Average Outcome as Target

Features:

- Expenditure data
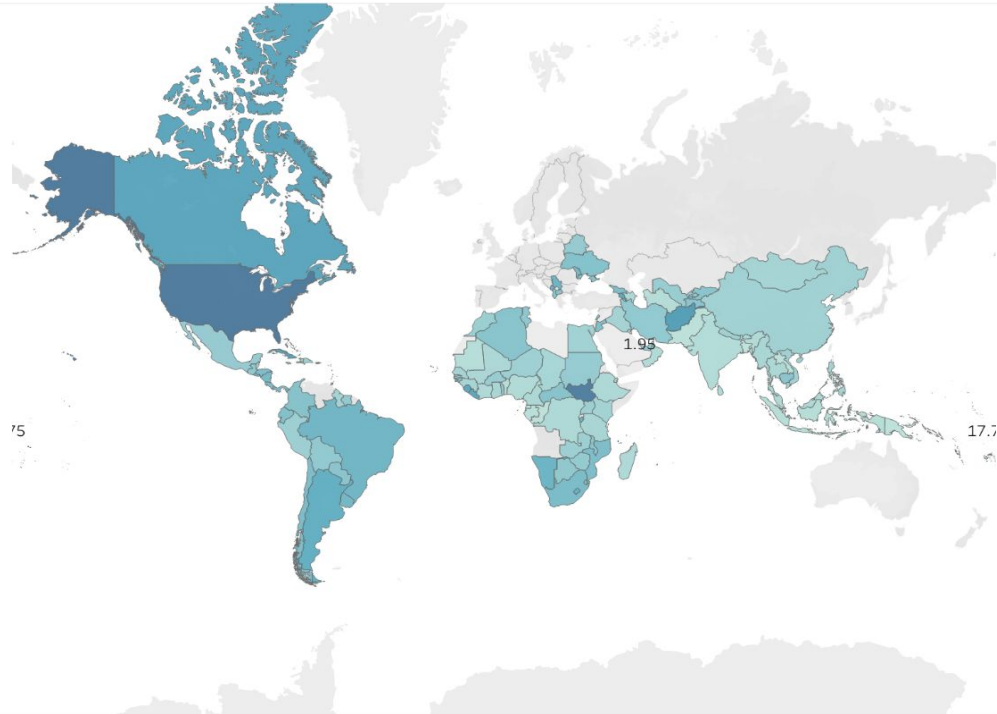- Excluding total number of deaths and labeling information

Global Health Expenditures

World Map
(Countries 2010-2022, Infectious Disease Expenditures as % of GDP)
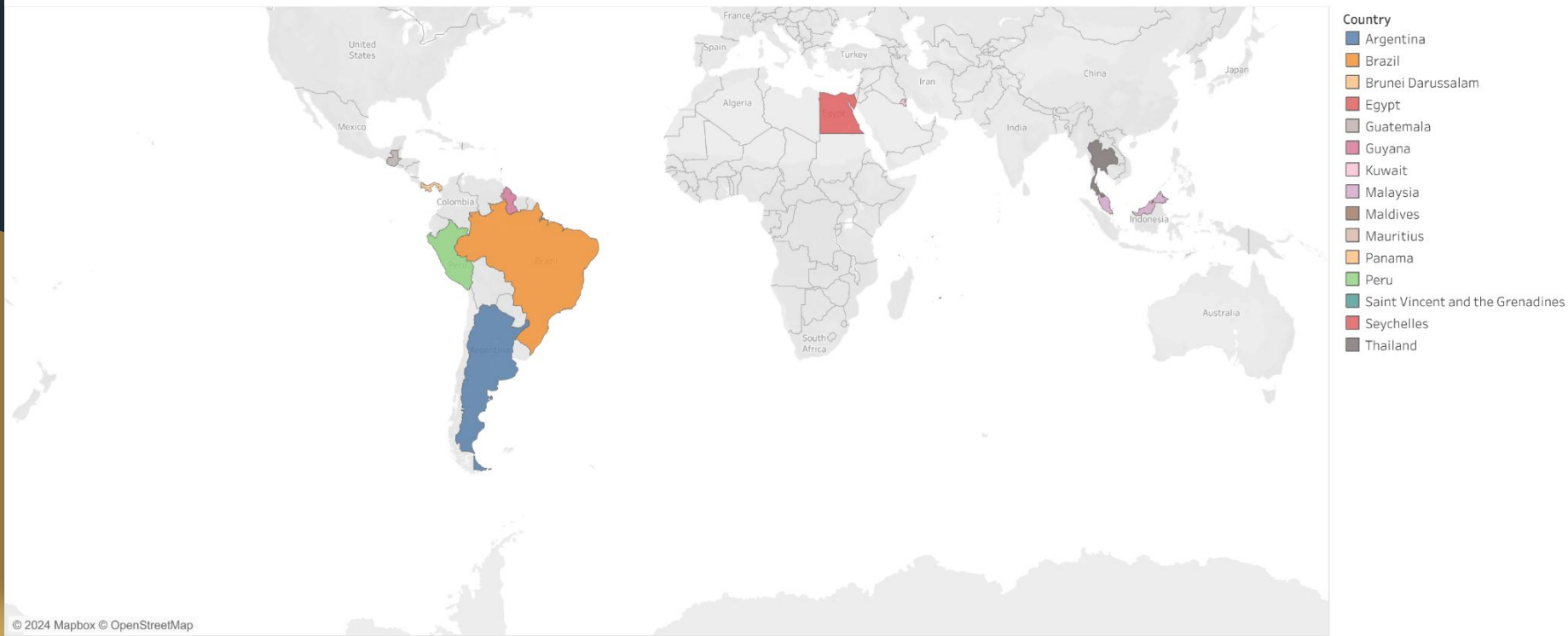Lowest: Qatar (1.95%)
Highest: Tuvalu (17.75%)

Looking Ahead: Countries with Mean Mortality Rate Greater Than Predicted by ML Model

"High Risk" of Mortality
(Countries with Current Mortality Rates Greater than Predictive Model Output)

Country
- Argentina
- Brazil
- Brunei Darussalam
- Egypt
- Guatemala
- Guyana
- Kuwait
- Malaysia
- Maldives
- Mauritius
- Panama
- Peru
- Saint Vincent and the Grenadines
- Seychelles
- Thailand

© 2024 Mapbox © OpenStreetMap

Strong, Steady Correlation Between Per Capita Health Expenditures And GDP (2010-2012)

Global Health Expenditures and GDP (per capita)
(Global Trends in Countries' Per Capita $ Spent on Infectious Diseases by Year, 2010-2022)
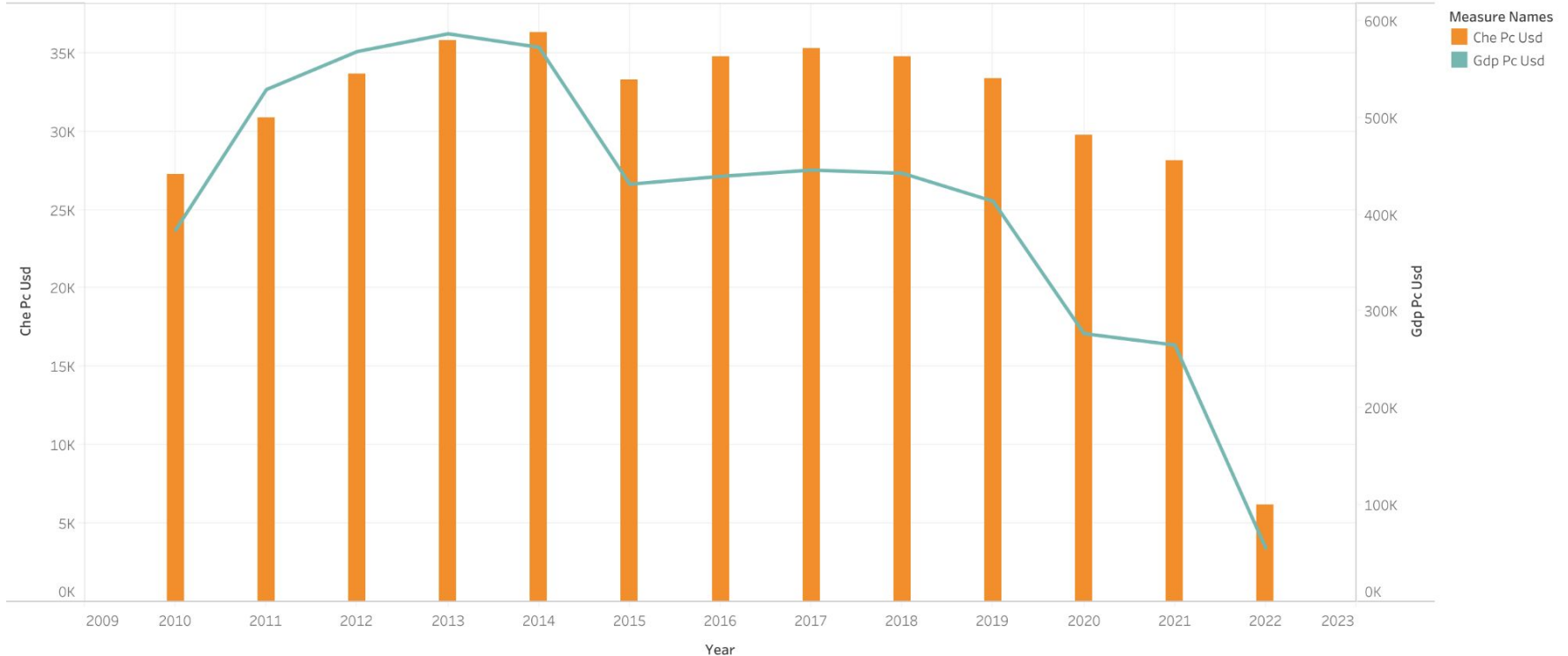
Higher Expenditures...Lower Mortality
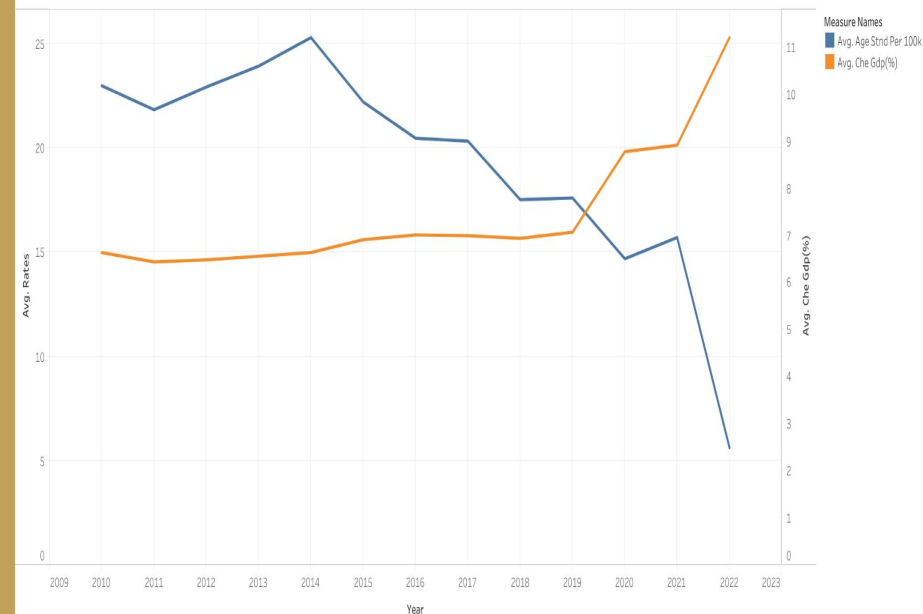## (Historical Spending and Mortality Trends, Average for all Countries))

Current Health Expenditures

Capital Health Expenditures
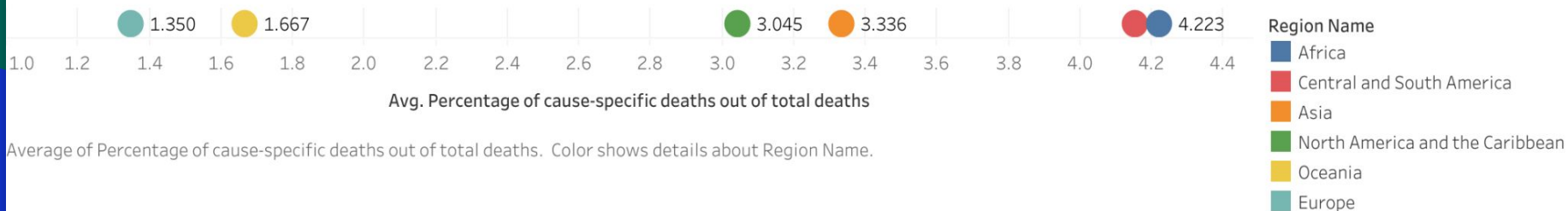

Average Mortality Rates vs. Capital Expenditure Rates Per Country (2010-2022)

Measure Names
— Avg. Age Stnd Per 100k
— Avg. Hk Gdl(%)
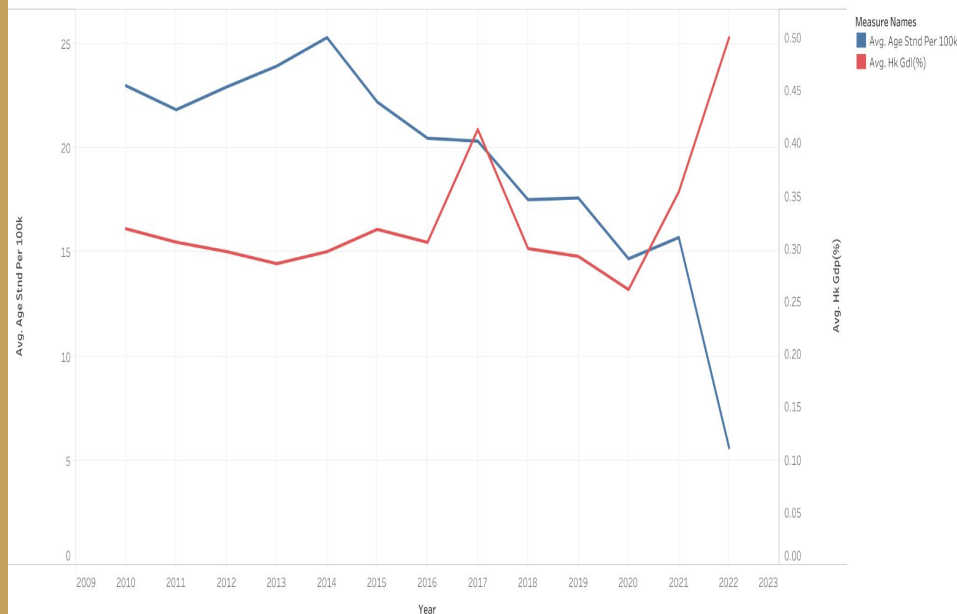
Average Country Mortality Rates vs. CHE Rates (2010-2022)

Measure Names
— Avg. Age Stnd Per 100k
— Avg. Che Gdp(%)

Africa, Central and South America are the most at-risk for Infectious Disease Mortality

Infectious Disease Specific Mortality Rates
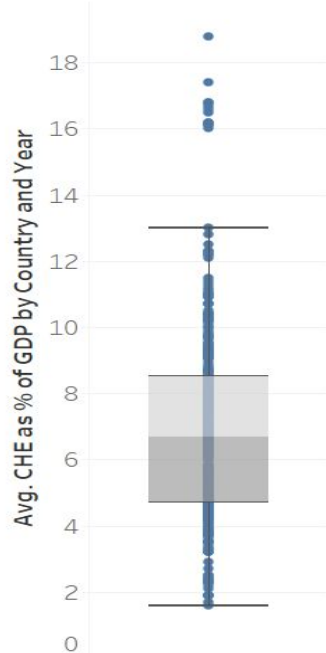## as Percentage of Total Mortality Rates
(2010-2022)

	1.350		1.667						3.045		3.336					4.223

1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4

Avg. Percentage of cause-specific deaths out of total deaths

Average of Percentage of cause-specific deaths out of total deaths. Color shows details about Region Name.

Region Name
- Africa
- Central and South America
- Asia
- North America and the Caribbean
- Oceania
- Europe

Identifying Outliers for Visualization
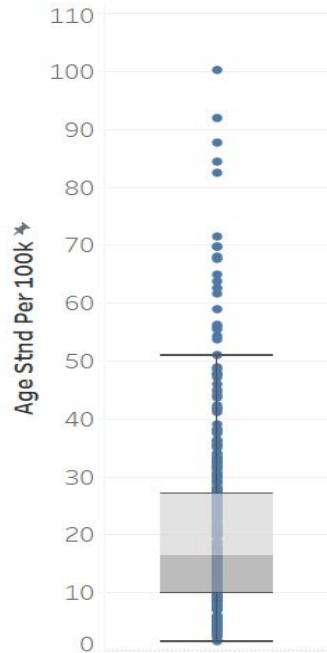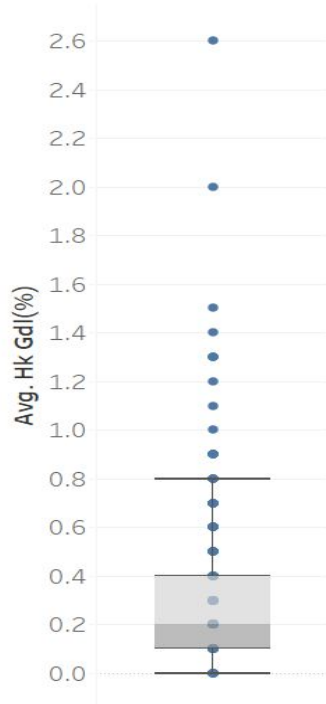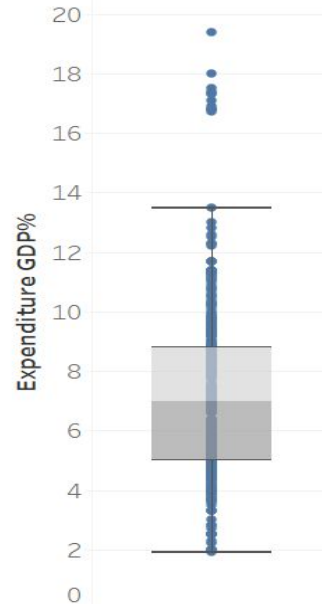
- Identified Main Columns for Linear Regressions
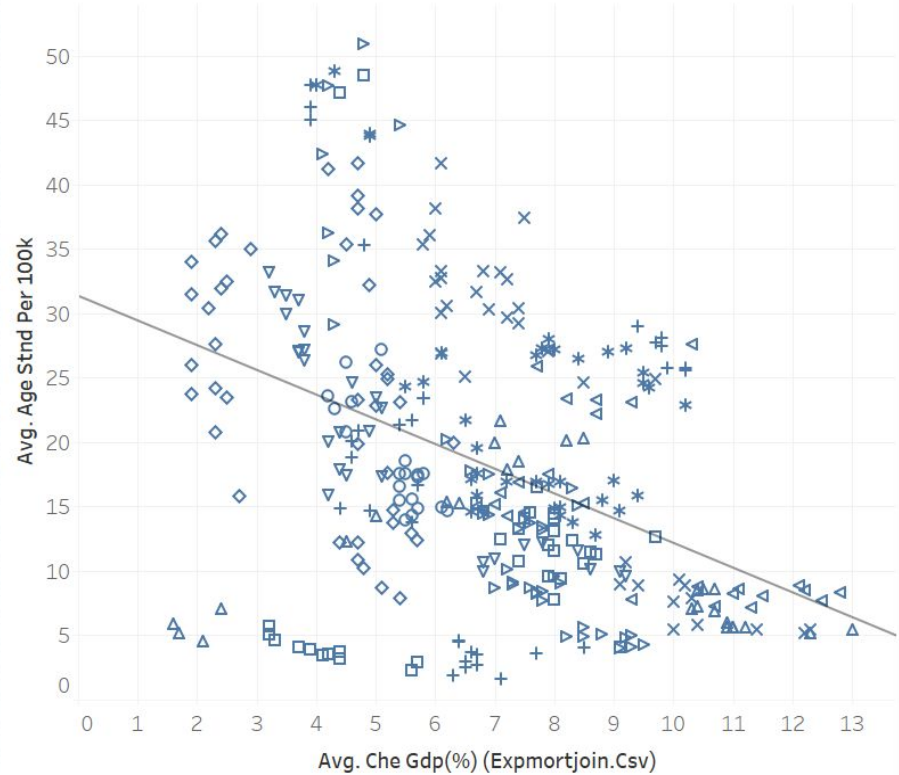- Used Tableau's Box Plots to Identify Outliers across metrics

R = -0.418 R² = 0.175 P-value < 0.0001   R = -0.42 R² =0.176 P-Value < 0.0001

CHE as % of GDP vs Age Stnd Per 100k (Outliers Included)

CHE as % of GDP vs Age Stnd Per 100k (Outliers Removed)

R = -0.041 R² = 0.0017 P-value =0.442   R = -0.0396 R² =0.0016 P-Value 0.466

HK as Pct of GDP vs Age Std Per 100k by Country and Year (Outliers Included)

HK as Pct of GDP vs Age Std Per 100k (Outliers Removed)
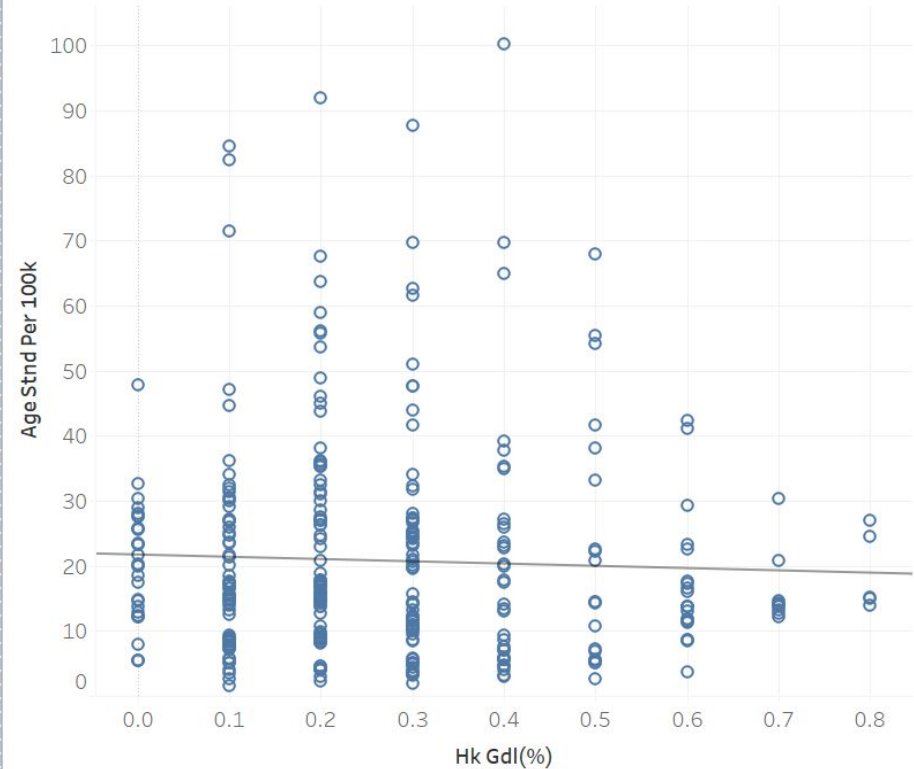
R = -0.406 R² = 0.165 P-value < 0.0001   R = -0.394 R² =0.155 P-Value < 0.0001

Agg Expenditure as % of GDP by Age Stnd Per 100k(Outliers Included)

Agg Expenditure as % of GDP by Age Stnd Per 100k(Outliers Removed)
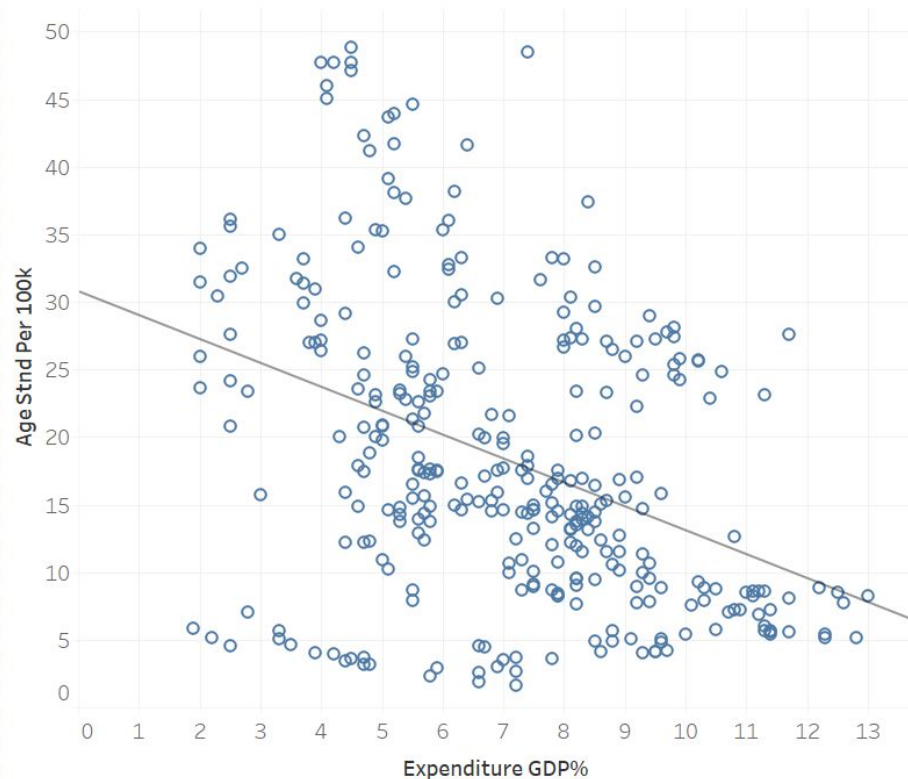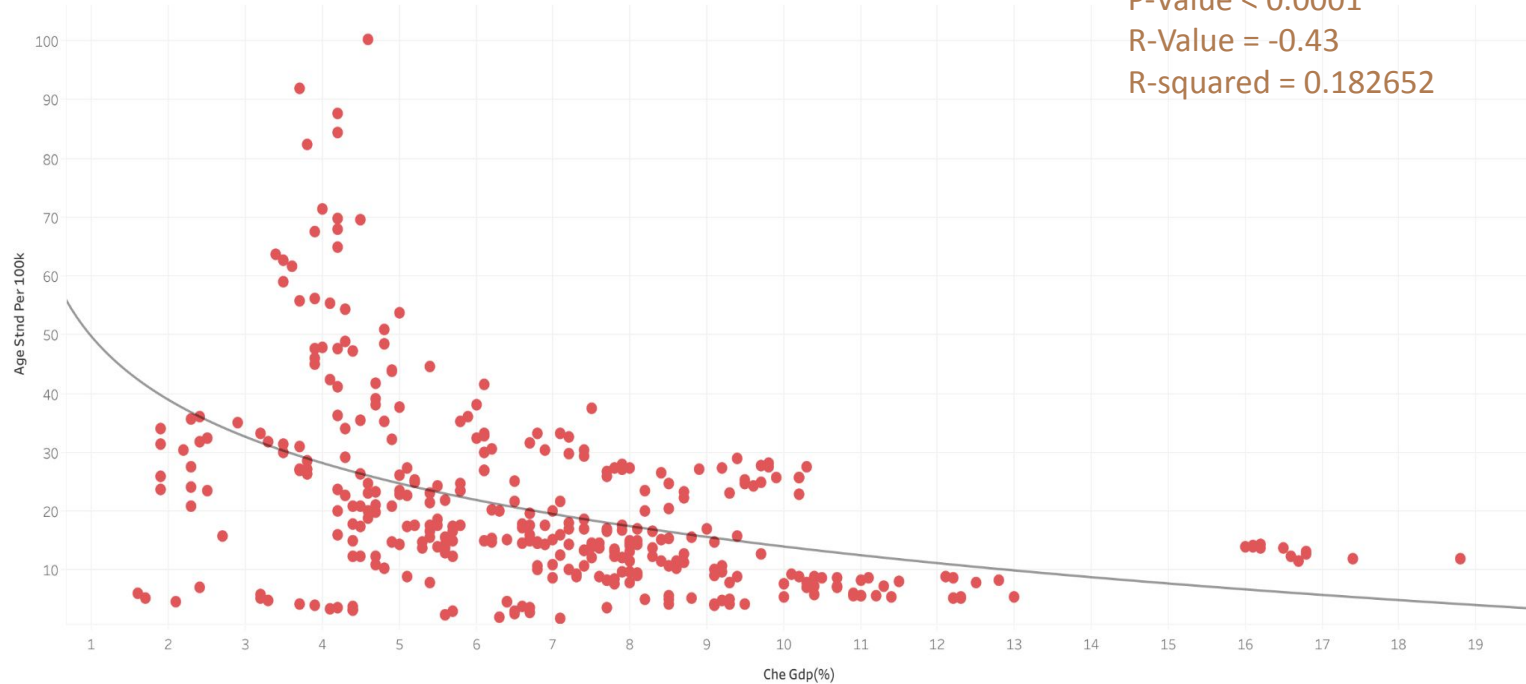
First Conclusion

Assumption of Linearity is incorrect, but there is some relationship of statistical significance.
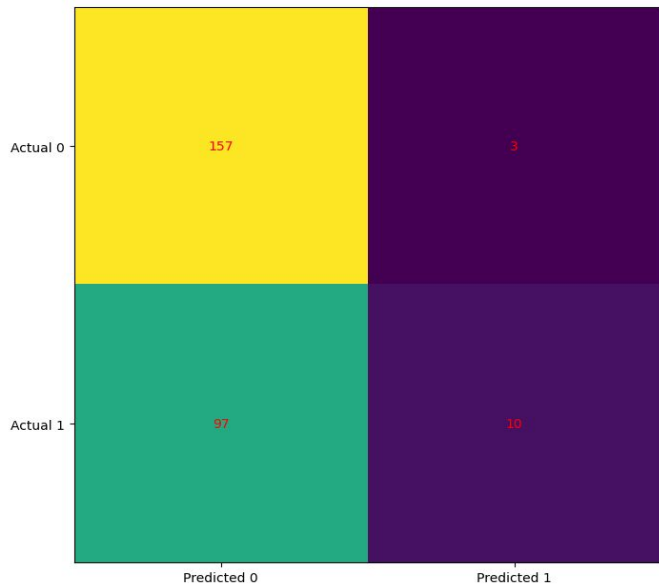
For Predictive Modeling: Utilize Logistic Regressions

Logistic Regression
(CHE % GDP vs Age Stnd per 100k, Outliers Included)

P-Value < 0.0001
R-Value = -0.43
R-squared = 0.182652

Logistic Regression

Training Data Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	157	3
Actual 1	97	10

Training Accuracy
63%

Testing Accuracy
67%

Testing Data Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	157	3
Actual 1	97	10
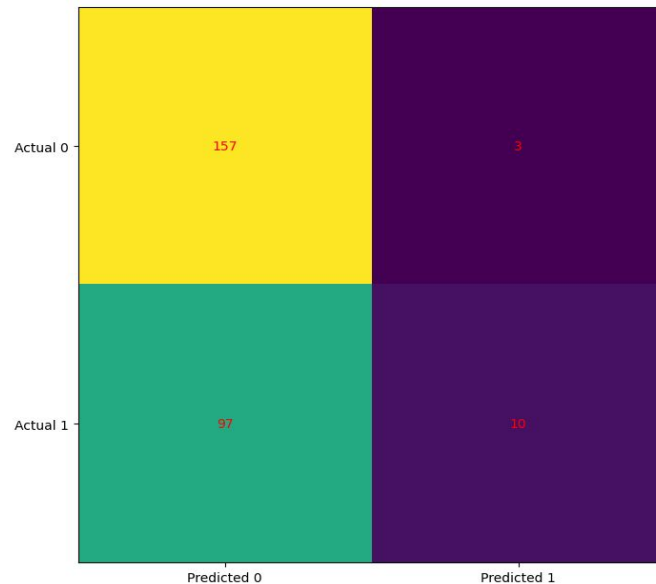
KNeighborsClassifier "Percent of Total"
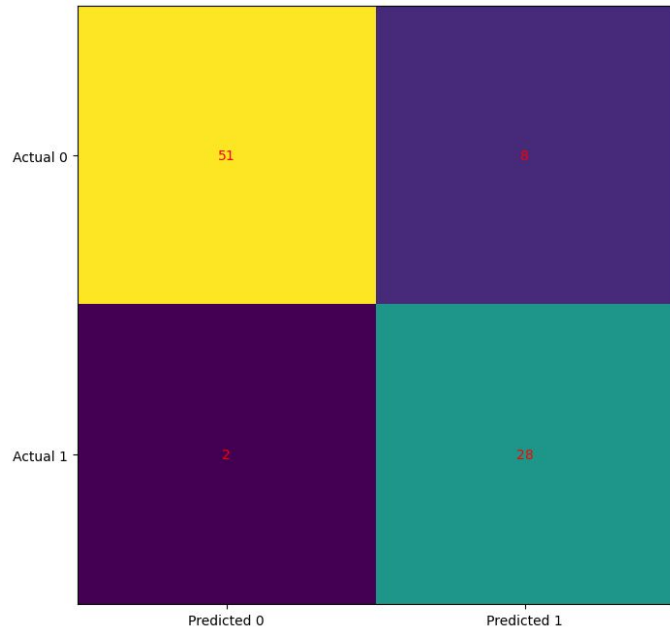
	precision	recall	f1-score	support
0	0.96	0.86	0.91	59
1	0.78	0.93	0.85	30
accuracy			0.89	

KNeighborsClassifier "Age Standardized"
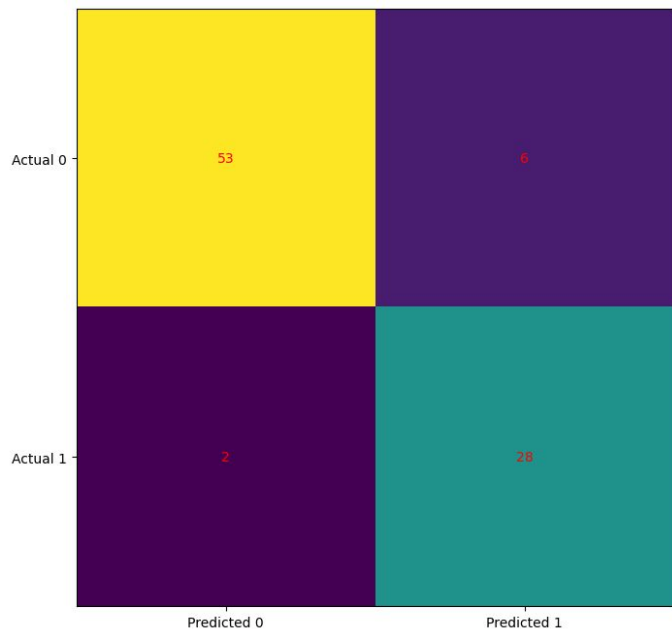
	precision	recall	f1-score	support
0	0.96	0.90	0.93	59
1	0.82	0.93	0.87	30
accuracy		0.91		

KNeighborsClassifier "Death Rate Per 100k"
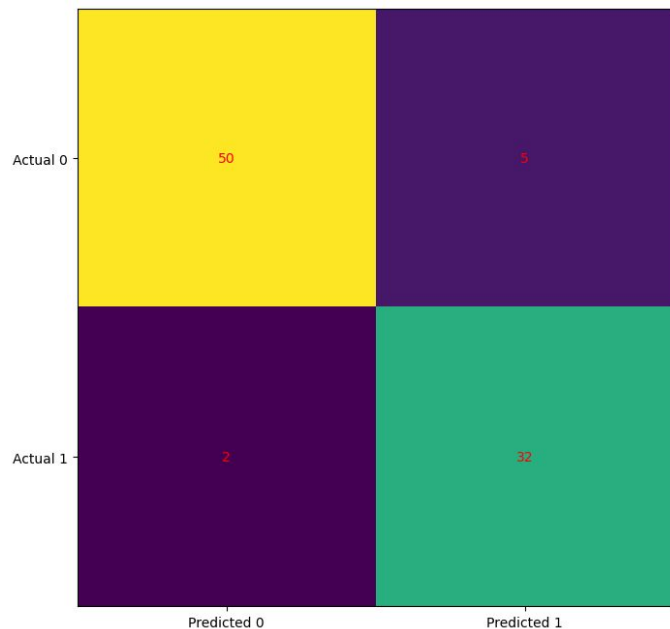
	precision	recall	f1-score	support
0	0.96	0.91	0.93	55
1	0.86	0.94	0.90	34
accuracy		0.92		

Model Conclusion

Key Takeaways

- Country-specific Health Expenditures are statistically significant in predicting mortality rates from infectious diseases.
- Current Health Expenditure(CHE) more statistically significant than Capital Health Expenditure (HK)
- Logistic regression, using a KKNeighbors Classifier, yields the most accurate predictive capabilities considering the complexity of our data.

Data Limitations

- Missing or incomplete health data
- Missing or incomplete expenditure data
- Missing context for cultural and socioeconomic variables
- Other significant covariables

Next Steps

- Explore WHO data more in depth
- Find other data sources
- Explore more non-linear predictive models

Questions?