TO VAX OR NOT? A SURVEY OF GLOBAL HEALTH INDICATORS

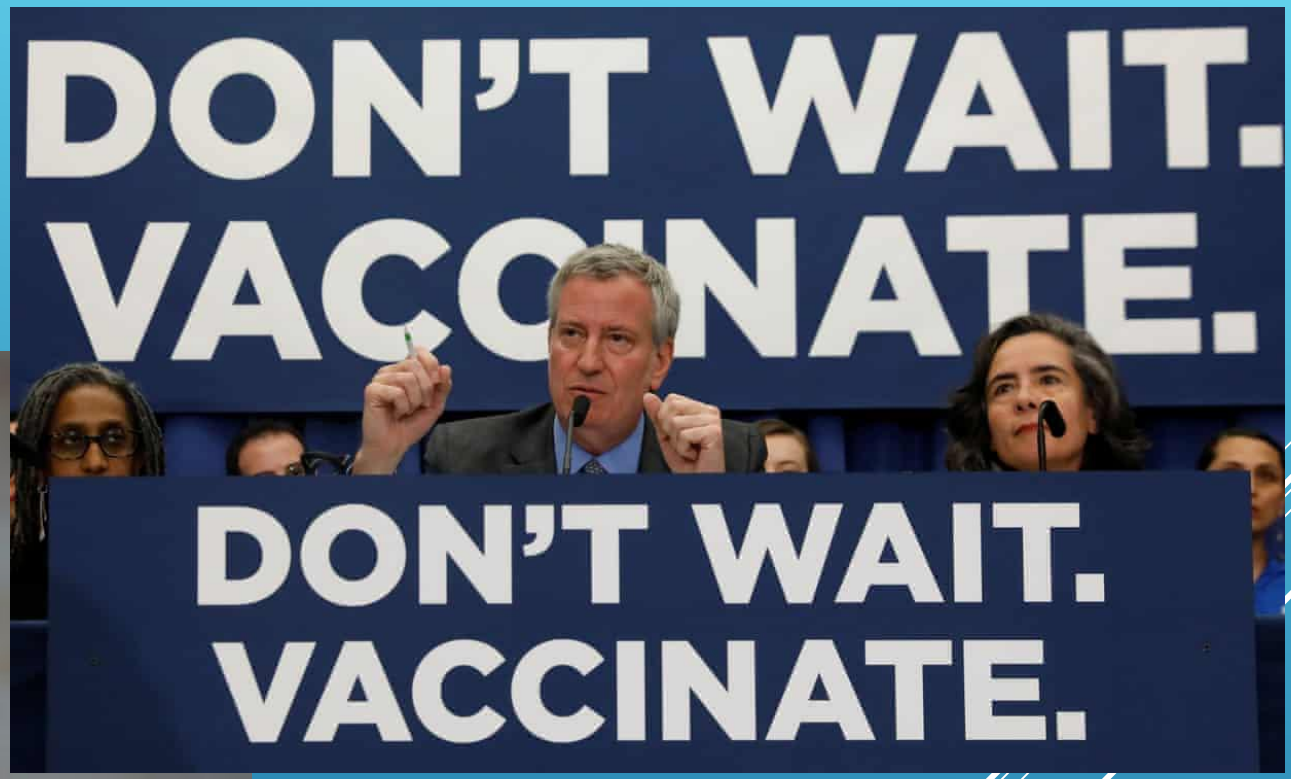
Shuja Sajjad

Kyaw Lynn

Jose Ramses Sanchez

THE CURRENT DEBATE

Ideology vs Science



Implications?

THE DATA HUNT

Data Tasks Kernels (30) Discussion (1) Activity Metadata Download (43 MB) New Notebook

Usability 5.9 Tags health, demographics, nutrition, vitamins

Description

Context

HealthStats provides key health, nutrition and population statistics gathered from a variety of international sources. Themes include population dynamics, nutrition, reproductive health, health financing, medical resources and usage, immunization, infectious diseases, HIV/AIDS, DALY, population projections and lending. HealthStats also includes health, nutrition and population statistics by wealth quintiles.

Content

This dataset includes 345 indicators, such as immunization rates, malnutrition prevalence, and vitamin A supplementation rates across 263 countries around the world. Data was collected on a yearly basis from 1960-2016.

- Sourced from the World Bank
- Over 300 various health indicators
- Data of 263 countries
- Historical data beginning from 1960 *(availability varied)*

SEARCH FOR A ~~CURE~~ DISEASE

Measles

- **Even though a safe and cost-effective vaccine is available, in 2018, there were more than 140,000 measles deaths globally. *(WHO)*[1]**
- **Most worldwide deaths occur to children under the age of five. *(WHO)*[1]**

[1]https://www.who.int/news-room/fact-sheets/detail/measles

QUESTIONS TO INVESTIGATE

- Is there a relationship between measles vaccination rates and infant mortality trends in the last decade?
- Also, do economic trends (measured by Gross National Product (GNP)) coincide with measles vaccination rates in the last decade?
- Countries to analyze: USA, China, Germany, and Brazil.

INITIAL DATA CLEANING

Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Arab World	ARB	% of females ages 15-49 having comprehensive corr	SH.HIV.KNOW.FE.ZS												
Arab World	ARB	% of males ages 15-49 having comprehensive correc	SH.HIV.KNOW.MA.ZS												
Arab World	ARB	Adolescent fertility rate (births per 1,000 women ag	SP.ADO.TFRT	133.555	134.1591	134.8579	134.5046	134.1052	133.5696	132.6756	131.6655	129.191	126.7368	124.3828	122.1334
Arab World	ARB	Adults (ages 15+) and children (0-14 years) living wit	SH.HIV.TOTL												
Arab World	ARB	Adults (ages 15+) and children (ages 0-14) newly inf	SH.HIV.INCD.TL												
Arab World	ARB	Adults (ages 15+) living with HIV	SH.DYN.AIDS												
Arab World	ARB	Adults (ages 15+) newly infected with HIV	SH.HIV.INCD												
Arab World	ARB	Age at first marriage, female	SP.DYN.SMAM.FE												
Arab World	ARB	Age at first marriage, male	SP.DYN.SMAM.MA												
Arab World	ARB	Age dependency ratio (% of working-age population	SP.POP.DPND	87.79923	89.22171	90.5024	91.60301	92.41492	92.86816	93.73395	94.10028	94.11993	93.96679	93.71193	93.9228
Arab World	ARB	Age dependency ratio, old	SP.POP.DPND.OL	6.635328	6.746978	6.84098	6.915871	6.968155	6.996499	7.065913	7.105005	7.122	7.12604	7.122081	7.163168
Arab World	ARB	Age dependency ratio, young	SP.POP.DPND.YG	81.02425	82.31614	83.45981	84.41397	85.09666	85.47072	86.24656	86.59086	86.63076	86.50189	86.25832	86.42596
Arab World	ARB	Age population, age 0, female, interpolated	SP.POP.AG00.FE.IN												
Arab World	ARB	Age population, age 0, male, interpolated	SP.POP.AG00.MA.IN												
Arab World	ARB	Age population, age 01, female, interpolated	SP.POP.AG01.FE.IN												
Arab World	ARB	Age population, age 01, male, interpolated	SP.POP.AG01.MA.IN												
Arab World	ARB	Age population, age 02, female, interpolated	SP.POP.AG02.FE.IN												
Arab World	ARB	Age population, age 02, male, interpolated	SP.POP.AG02.MA.IN												
Arab World	ARB	Age population, age 03, female, interpolated	SP.POP.AG03.FE.IN												
Arab World	ARB	Age population, age 03, male, interpolated	SP.POP.AG03.MA.IN												
Arab World	ARB	Age population, age 04, female, interpolated	SP.POP.AG04.FE.IN												
Arab World	ARB	Age population, age 04, male, interpolated	SP.POP.AG04.MA.IN												
Arab World	ARB	Age population, age 05, female, interpolated	SP.POP.AG05.FE.IN												
Arab World	ARB	Age population, age 05, male, interpolated	SP.POP.AG05.MA.IN												
Arab World	ARB	Age population, age 06, female, interpolated	SP.POP.AG06.FE.IN												
Arab World	ARB	Age population, age 06, male, interpolated	SP.POP.AG06.MA.IN												
Arab World	ARB	Age population, age 07, female, interpolated	SP.POP.AG07.FE.IN												
Arab World	ARB	Age population, age 07, male, interpolated	SP.POP.AG07.MA.IN												
Arab World	ARB	Age population, age 08, female, interpolated	SP.POP.AG08.FE.IN												
Arab World	ARB	Age population, age 08, male, interpolated	SP.POP.AG08.MA.IN												
Arab World	ARB	Age population, age 09, female, interpolated	SP.POP.AG09.FE.IN												
Arab World	ARB	Age population, age 09, male, interpolated	SP.POP.AG09.MA.IN												
Arab World	ARB	Age population, age 10, female, interpolated	SP.POP.AG10.FE.IN												
Arab World	ARB	Age population, age 10, male, interpolated	SP.POP.AG10.MA.IN												
Arab World	ARB	Age population, age 11, female, interpolated	SP.POP.AG11.FE.IN												
Arab World	ARB	Age population, age 11, male, interpolated	SP.POP.AG11.MA.IN												

INITIAL DATA CLEANING (USA)

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import json
from scipy.stats import linregress
from scipy import stats

yr_list= [2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013,
       2014, 2015]

# Hide warning messages in notebook
import warnings
warnings.filterwarnings('ignore')

# File to Load
data_to_load = "data - testing2.csv"

# Read the Population Health Data
health_data_pd = pd.read_csv(data_to_load)

# Display the data table for preview
health_data_pd
```

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
0	Arab World	ARB	% of females ages 15-49 having comprehensive c...	SH.HIV.KNOW.FE.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	% of males ages 15-49 having comprehensive cor...	SH.HIV.KNOW.MA.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adolescent fertility rate (births per 1,000 wo...	SP.ADO.TFRT	133.555013	134.159119	134.857912	134.504576	134.105211	133.569626	...	50.329135	49.999851	49.887046	49.781207	49.672975	49.536047	49.383745	48.796558	48.196418	NaN
3	Arab World	ARB	Adults (ages 15+) and children (0-14 years) li...	SH.HIV.TOTL	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Arab World	ARB	Adults (ages 15+) and children (ages 0-14) new...	SH.HIV.INCD.TL	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89005	Zimbabwe	ZWE	Use of insecticide-treated bed nets (% of unde...	SH.MLR.NETS.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	3.100000	NaN	NaN	17.300000	NaN	9.700000	NaN	NaN	26.800000	NaN
89006	Zimbabwe	ZWE	Use of Intermittent Preventive Treatment of ma...	SH.MLR.SPF2.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	6.300000	NaN	NaN	13.900000	NaN	7.300000	NaN	NaN	12.900000	NaN
89007	Zimbabwe	ZWE	Vitamin A supplementation coverage rate (% of ...	SN.ITK.VITA.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	67.000000	83.000000	0.000000	77.000000	49.000000	47.000000	61.000000	34.000000	32.000000	NaN
89008	Zimbabwe	ZWE	Wanted fertility rate (births per woman)	SP.DYN.WFRT	NaN	NaN	NaN	NaN	NaN	NaN	...	3.300000	NaN	NaN	NaN	NaN	3.500000	NaN	NaN	NaN	NaN
89009	Zimbabwe	ZWE	Women's share of population ages 15+ living wi...	SH.DYN.AIDS.FE.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	58.331272	58.586086	58.760796	58.812421	58.825943	58.899308	58.939080	58.900126	58.822335	58.855552

89010 rows × 60 columns

INITIAL DATA CLEANING (USA)

```
health_data_decade_df = health_data_pd[['Country Name','Country Code','Indicator Name','Indicator Code',
                                        '2005','2006','2007','2008','2009','2010','2011','2012','2013','2014','2015']]
health_data_decade_df
```

	Country Name	Country Code	Indicator Name	Indicator Code	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
0	Arab World	ARB	% of females ages 15-49 having comprehensive c...	SH.HIV.KNOW.FE.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	% of males ages 15-49 having comprehensive cor...	SH.HIV.KNOW.MA.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adolescent fertility rate (births per 1,000 wo...	SP.ADO.TFRT	50.732590	50.329135	49.999851	49.887046	49.781207	49.672975	49.536047	49.383745	48.796558	48.196418	NaN
3	Arab World	ARB	Adults (ages 15+) and children (0-14 years) li...	SH.HIV.TOTL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Arab World	ARB	Adults (ages 15+) and children (ages 0-14) new...	SH.HIV.INCD.TL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89005	Zimbabwe	ZWE	Use of insecticide-treated bed nets (% of unde...	SH.MLR.NETS.ZS	NaN	3.100000	NaN	NaN	17.300000	NaN	9.700000	NaN	NaN	26.800000	NaN
89006	Zimbabwe	ZWE	Use of Intermittent Preventive Treatment of ma...	SH.MLR.SPF2.ZS	NaN	6.300000	NaN	NaN	13.900000	NaN	7.300000	NaN	NaN	12.900000	NaN
89007	Zimbabwe	ZWE	Vitamin A supplementation coverage rate (% of ...	SN.ITK.VITA.ZS	81.000000	67.000000	83.000000	0.000000	77.000000	49.000000	47.000000	61.000000	34.000000	32.000000	NaN
89008	Zimbabwe	ZWE	Wanted fertility rate (births per woman)	SP.DYN.WFRT	NaN	3.300000	NaN	NaN	NaN	NaN	3.500000	NaN	NaN	NaN	NaN
89009	Zimbabwe	ZWE	Women's share of population ages 15+ living wi...	SH.DYN.AIDS.FE.ZS	58.045596	58.331272	58.586086	58.760796	58.812421	58.825943	58.899308	58.939080	58.900126	58.822335	58.855552

89010 rows × 15 columns

INITIAL DATA CLEANING (USA)

```
countries_sorted_df = health_data_decade_df.groupby('Country Name')
countries_sorted_df
usa_df = countries_sorted_df.get_group('United States')
usa_df
```

	Country Name	Country Code	Indicator Name	Indicator Code	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
85215	United States	USA	% of females ages 15-49 having comprehensive c...	SH.HIV.KNOW.FE.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85216	United States	USA	% of males ages 15-49 having comprehensive cor...	SH.HIV.KNOW.MA.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85217	United States	USA	Adolescent fertility rate (births per 1,000 wo...	SP.ADO.TFRT	41.0818	40.3754	39.669	37.7398	35.8106	33.8814	31.9522	30.023	27.0666	24.1102	NaN
85218	United States	USA	Adults (ages 15+) and children (0-14 years) li...	SH.HIV.TOTL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85219	United States	USA	Adults (ages 15+) and children (ages 0-14) new...	SH.HIV.INCD.TL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
85555	United States	USA	Use of insecticide-treated bed nets (% of unde...	SH.MLR.NETS.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85556	United States	USA	Use of Intermittent Preventive Treatment of ma...	SH.MLR.SPF2.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85557	United States	USA	Vitamin A supplementation coverage rate (% of ...	SN.ITK.VITA.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85558	United States	USA	Wanted fertility rate (births per woman)	SP.DYN.WFRT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
85559	United States	USA	Women's share of population ages 15+ living wi...	SH.DYN.AIDS.FE.ZS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

345 rows × 15 columns

INITIAL DATA CLEANING (USA)

```
dropped_usa_df = usa_df.dropna()
dropped_usa_df
```

	Country Name	Country Code	Indicator Name	Indicator Code	2005	2006	2007	2008	2009	2010	2011	2012	
85224	United States	USA	Age dependency ratio (% of working-age populat...	SP.POP.DPND	4.869685e+01	4.871720e+01	4.864902e+01	4.857000e+01	4.858461e+01	4.873401e+01	4.891222e+01	4.921263e+01	4.967
85225	United States	USA	Age dependency ratio, old	SP.POP.DPND.OL	1.832077e+01	1.842681e+01	1.855893e+01	1.874134e+01	1.900255e+01	1.935391e+01	1.979159e+01	2.032804e+01	2.094
85226	United States	USA	Age dependency ratio, young	SP.POP.DPND.YG	3.037608e+01	3.029038e+01	3.009009e+01	2.982867e+01	2.958206e+01	2.938010e+01	2.912063e+01	2.888459e+01	2.872
85227	United States	USA	Age population, age 0, female, interpolated	SP.POP.AG00.FE.IN	2.074479e+06	2.082882e+06	2.070141e+06	2.041092e+06	2.003862e+06	1.965540e+06	1.928888e+06	1.890396e+06	1.858
85228	United States	USA	Age population, age 0, male, interpolated	SP.POP.AG00.MA.IN	2.158649e+06	2.167520e+06	2.155450e+06	2.127284e+06	2.091058e+06	2.053711e+06	2.016453e+06	1.977553e+06	1.945
...	...	...	...	...	...	...	...	...	...	...	...	...	
85518	United States	USA	Rural population (% of total population)	SP.RUR.TOTL.ZS	2.007200e+01	1.990100e+01	1.973100e+01	1.956200e+01	1.939400e+01	1.922800e+01	1.906000e+01	1.889200e+01	1.872
85519	United States	USA	Rural population growth (annual %)	SP.RUR.TOTL.ZG	7.338981e-02	1.086703e-01	9.315709e-02	8.565726e-02	1.413321e-02	-2.362652e-02	-1.137156e-01	-1.235261e-01	-1.611
85550	United States	USA	Urban population	SP.URB.TOTL	2.362005e+08	2.389993e+08	2.417953e+08	2.446071e+08	2.472763e+08	2.498656e+08	2.523052e+08	2.547624e+08	2.571
85551	United States	USA	Urban population (% of total)	SP.URB.TOTL.IN.ZS	7.992800e+01	8.009900e+01	8.026900e+01	8.043800e+01	8.060600e+01	8.077200e+01	8.094000e+01	8.110800e+01	8.127
85552	United States	USA	Urban population growth (annual %)	SP.URB.GROW	1.135885e+00	1.177968e+00	1.163068e+00	1.156186e+00	1.085290e+00	1.041720e+00	9.716271e-01	9.691533e-01	9.455

175 rows × 15 columns

INITIAL DATA CLEANING (USA)

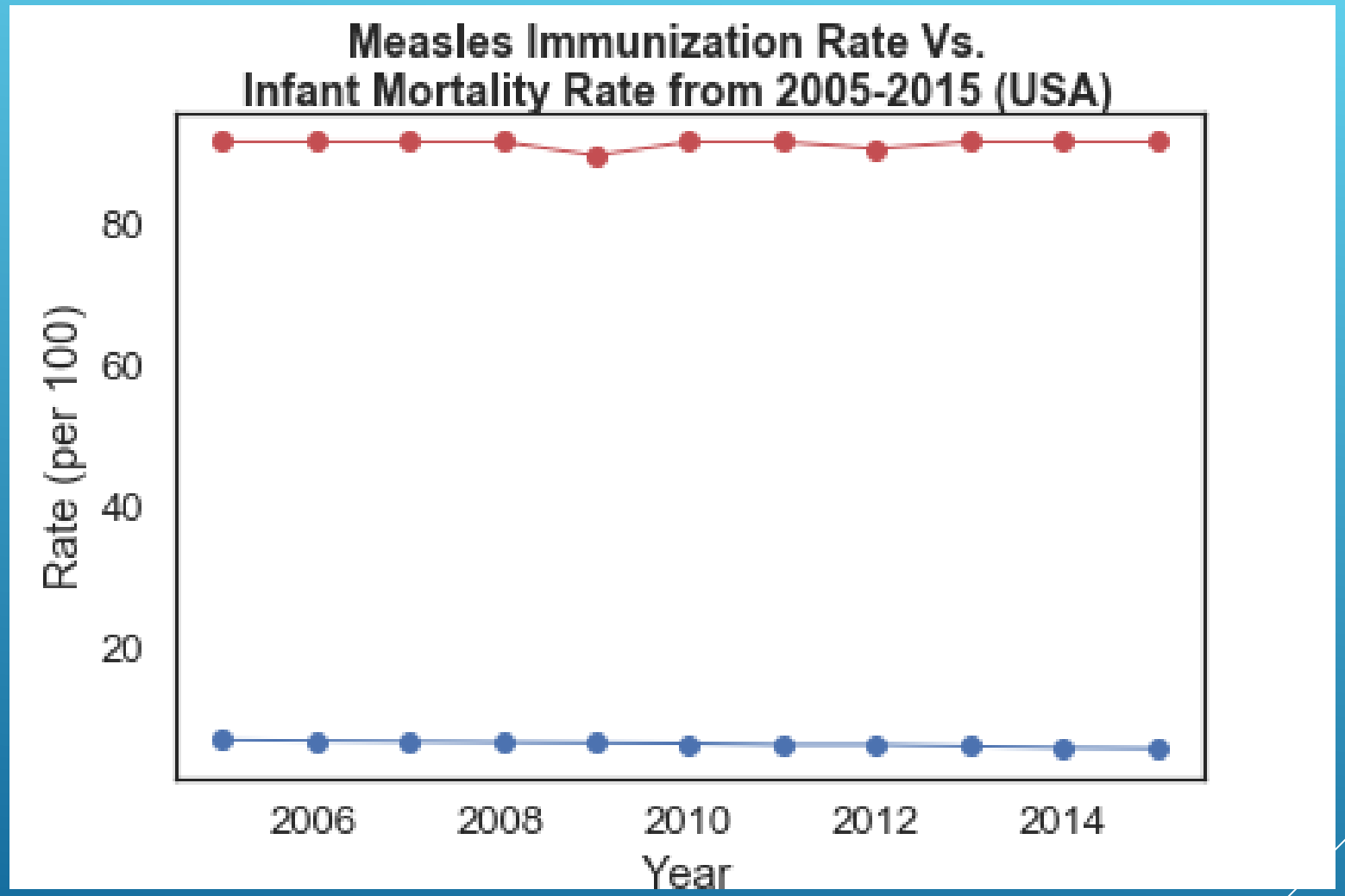
```
decade_health_USA_T = dropped_usa_df.iloc[:, 2:].T
decade_health_USA_T.columns = decade_health_USA_T.iloc[1,:]
decade_health_USA_T = decade_health_USA_T.iloc[2:, :]

decade_health_USA_T_codes = decade_health_USA_T[['SH.IMM.HEPB', 'SH.IMM.MEAS', 'SH.STA.ACSN',
                                                 'SP.DYN.IMRT.IN','SH.H2O.SAFE.ZS','SP.POP.GROW','SP.POP.TOTL','NY.GNP.PCAP.CD']]
decade_health_USA_T_codes
```

Indicator Code	SH.IMM.HEPB	SH.IMM.MEAS	SH.STA.ACSN	SP.DYN.IMRT.IN	SH.H2O.SAFE.ZS	SP.POP.GROW	SP.POP.TOTL	NY.GNP.PCAP.CD
2005	93	92	99.8	6.8	99	0.921713	2.95517e+08	46340
2006	93	92	99.9	6.7	99	0.964254	2.9838e+08	48080
2007	93	92	99.9	6.6	99	0.951055	3.01231e+08	48640
2008	94	92	99.9	6.5	99	0.945865	3.04094e+08	49330
2009	92	90	99.9	6.4	99.1	0.876651	3.06772e+08	48050
2010	92	92	99.9	6.3	99.1	0.835992	3.09347e+08	48950
2011	91	92	100	6.1	99.1	0.76385	3.11719e+08	50450
2012	90	91	100	6.1	99.1	0.761808	3.14103e+08	52520
2013	91	92	100	5.9	99.2	0.737406	3.16427e+08	53670
2014	92	92	100	5.7	99.2	0.780697	3.18907e+08	54400
2015	92	92	100	5.6	99.2	0.784423	3.21419e+08	54960

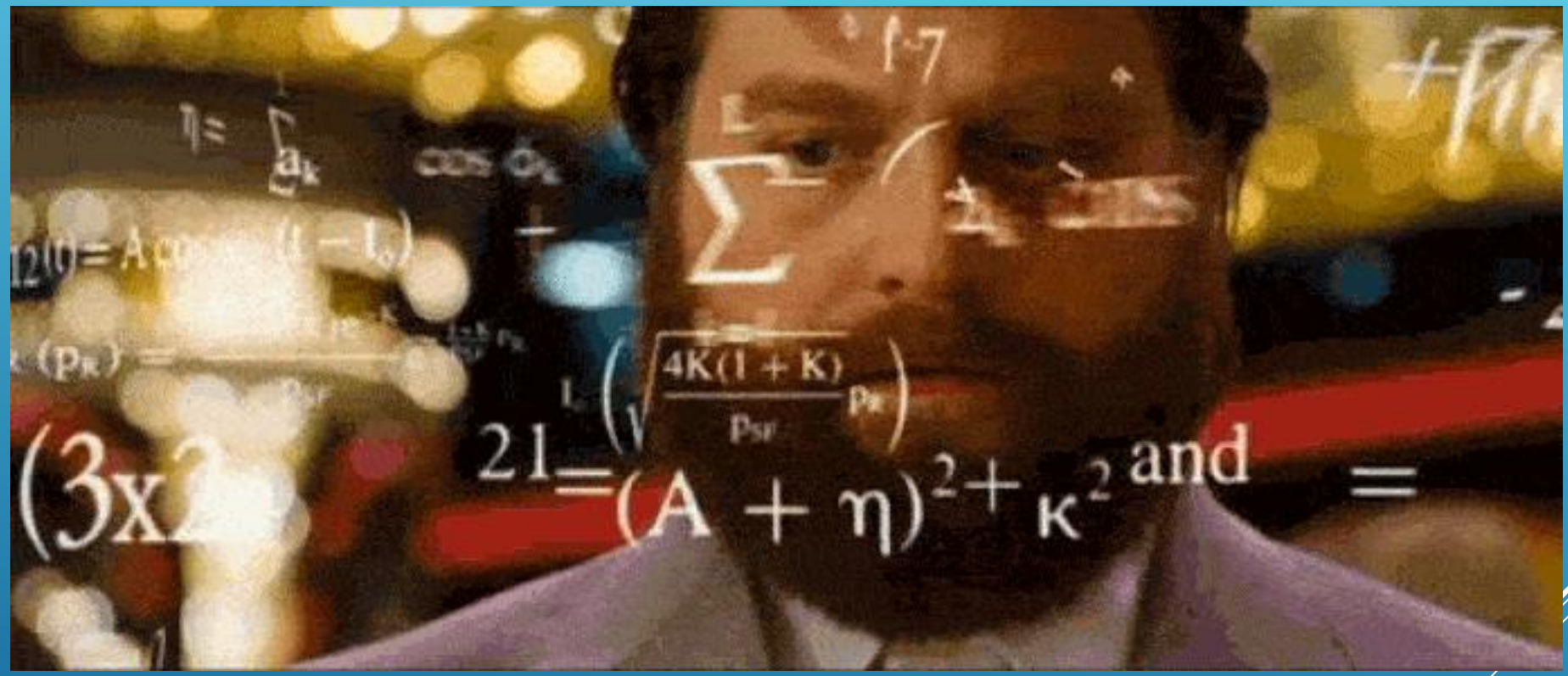
INITIAL DATA CLEANING (USA)

```
fig, (ax1) = plt.subplots(1, sharex=True)
fig.suptitle('Measles Immunization Rate Vs. \n Infant Mortality Rate from 2005-2015 (USA)', fontsize=14, fontweight="bold")

yr_list= [2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013,
       2014, 2015]

x = yr_list
y = [ x[0] for x in decade_health_USA_T_codes[['SP.DYN.IMRT.IN']].values]

ax1.set_xlim(min(yr_list)-.5, max(yr_list)+.5)
ax1.plot(x, y, linewidth=1, marker="o")

x = yr_list
y = [ x[0] for x in decade_health_USA_T_codes[['SH.IMM.MEAS']].values]

ax1.plot(x, y, linewidth=1, marker="o", color="r")
ax1.set_ylabel("Rate (per 100)")
ax1.set_xlabel("Year")

plt.savefig("measles_usa.png")
```

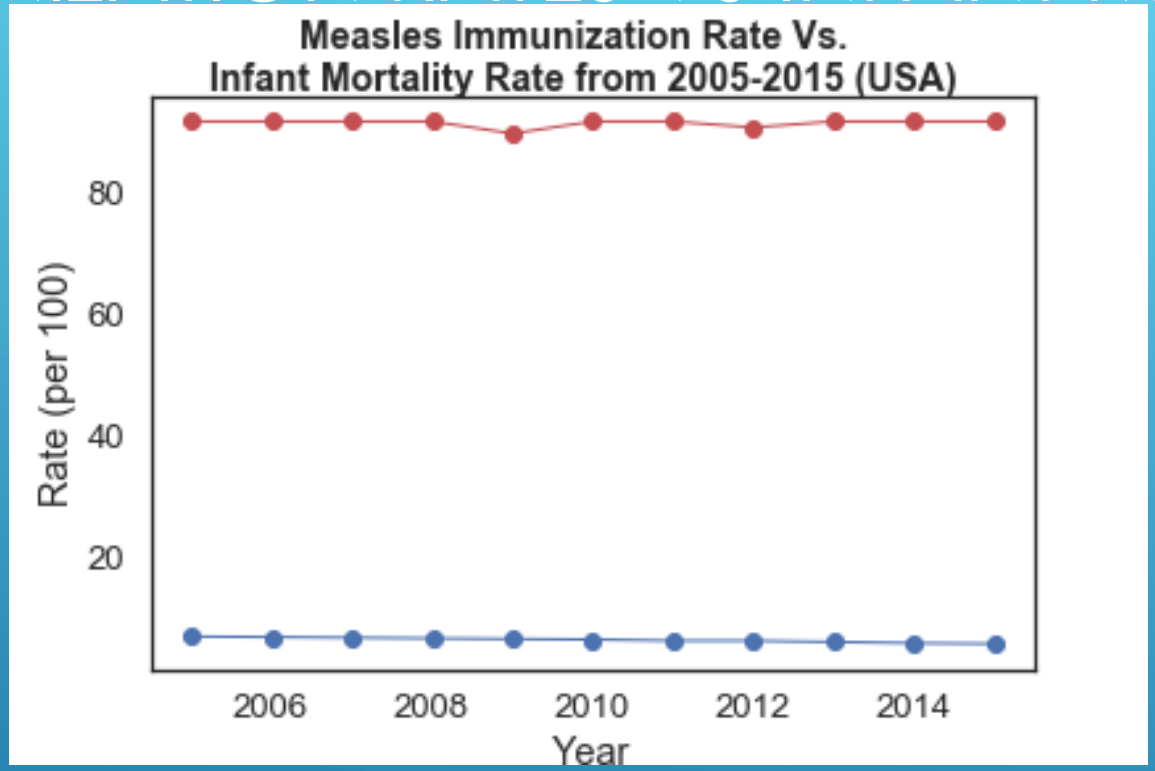
INITIAL DATA CLEANING (USA)



DATA ANALYSIS



DATA ANALYSIS
IMMUNIZATION RATES VS INFANT MORTALITY




```
import pingouin as pg

x = [ x[0] for x in decade_health_USA_T_codes[['SP.DYN.IMRT.IN']].values]
y = [ x[0] for x in decade_health_USA_T_codes[['SH.IMM.MEAS']].values]

pg.corr(x, y)
```

	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	11	-0.063	[-0.64, 0.56]	0.004	-0.245	0.853729	0.375	0.053

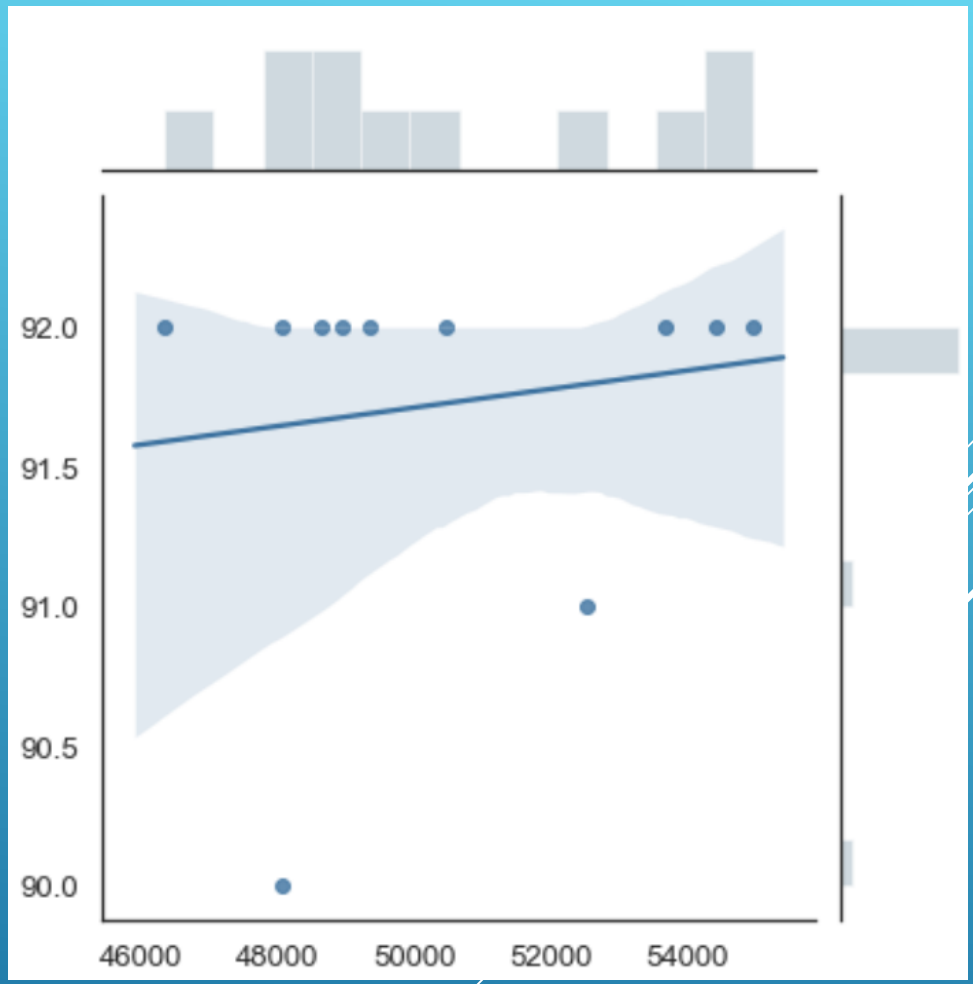
Note negative r value

DATA ANALYSIS
GNP AND IMMUNIZATION RATES

```
import pingouin as pg
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
sns.set(style='white', font_scale=1.2)

x = [ x[0] for x in decade_health_USA_T_codes[['NY.GNP.PCAP.CD']].values]
y = [ x[0] for x in decade_health_USA_T_codes[['SH.IMM.MEAS']].values]

print(pg.corr(x, y))

g = sns.JointGrid(x, y)
g = g.plot_joint(sns.regplot, color="xkcd:muted blue")
g = g.plot_marginals(sns.distplot, kde=False, bins=12, color="xkcd:bluey grey")
g.ax_joint.text(150, 95, 'r = 0.45, p < .001', fontstyle='italic')
plt.tight_layout()
```

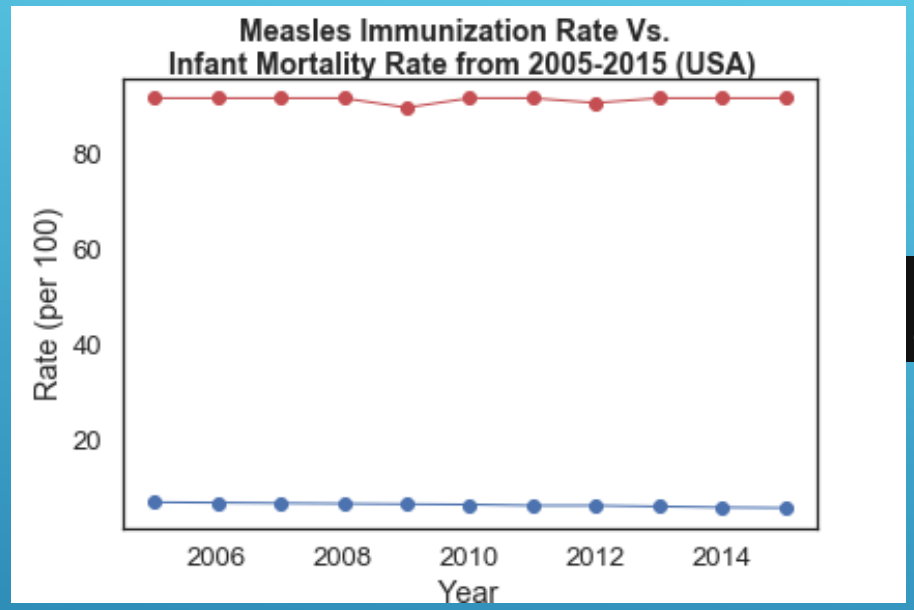
	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	11	0.151	[-0.49, 0.69]	0.023	-0.222	0.658471	0.403	0.072

Note positive r value

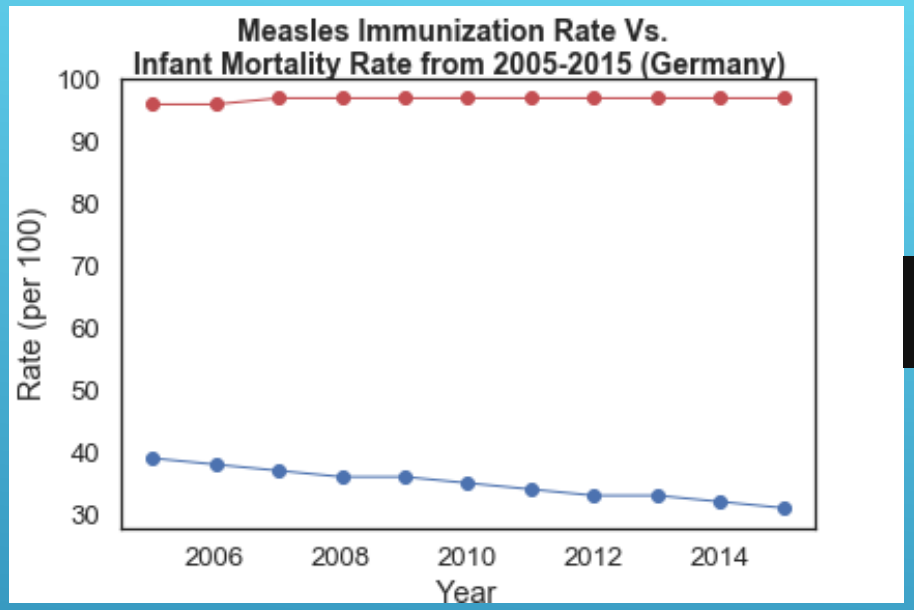


X axis – GNP (in millions of $)
Y axis – Measles Immunization Rate

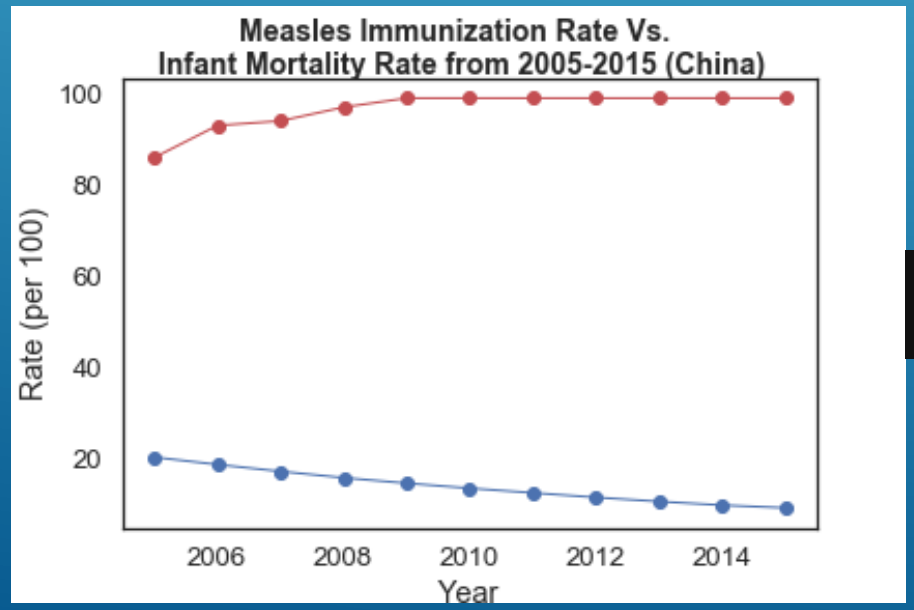
IMMUNIZATION AND INFANT MORTALITY CORRELATION



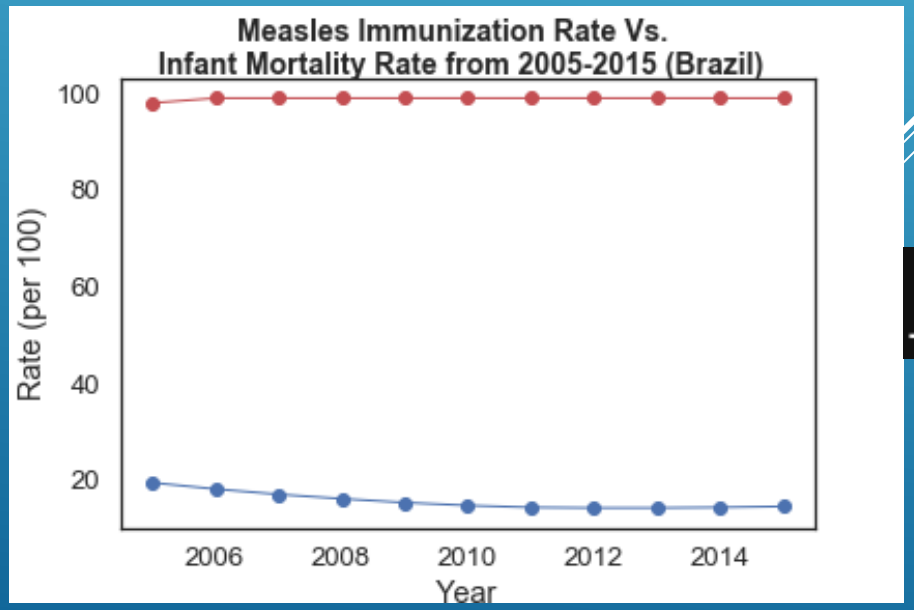
r
-0.063



r
-0.697



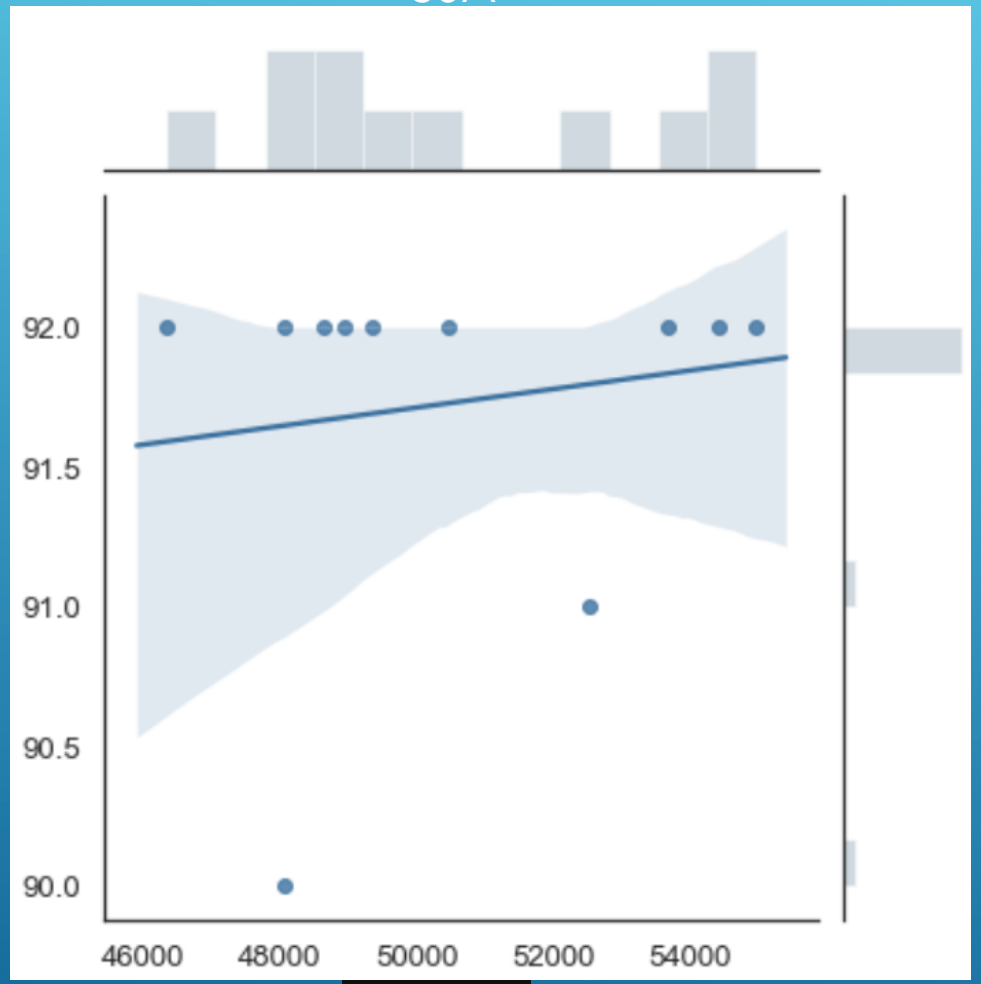
r
-0.849



r
-0.693

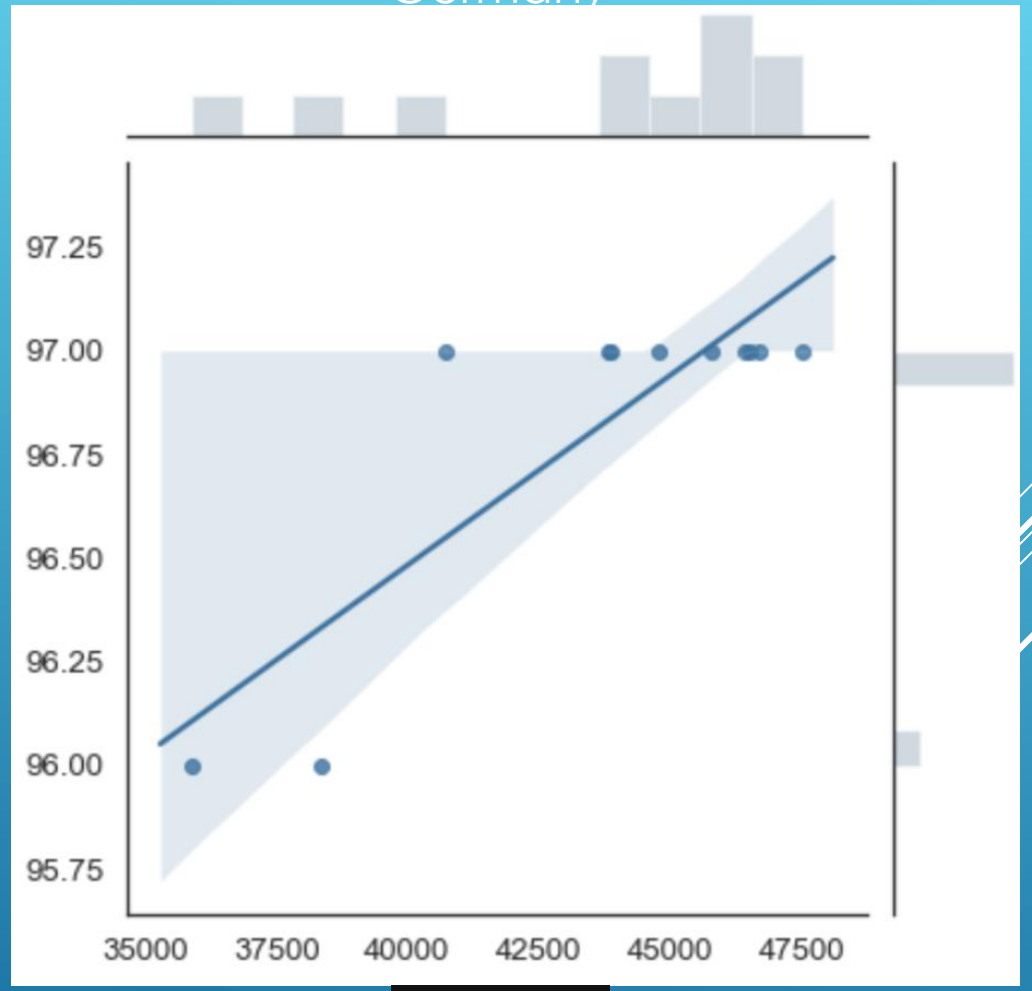
GNP AND IMMUNIZATION RATE CORRELATION

USA



r
0.151

Germany

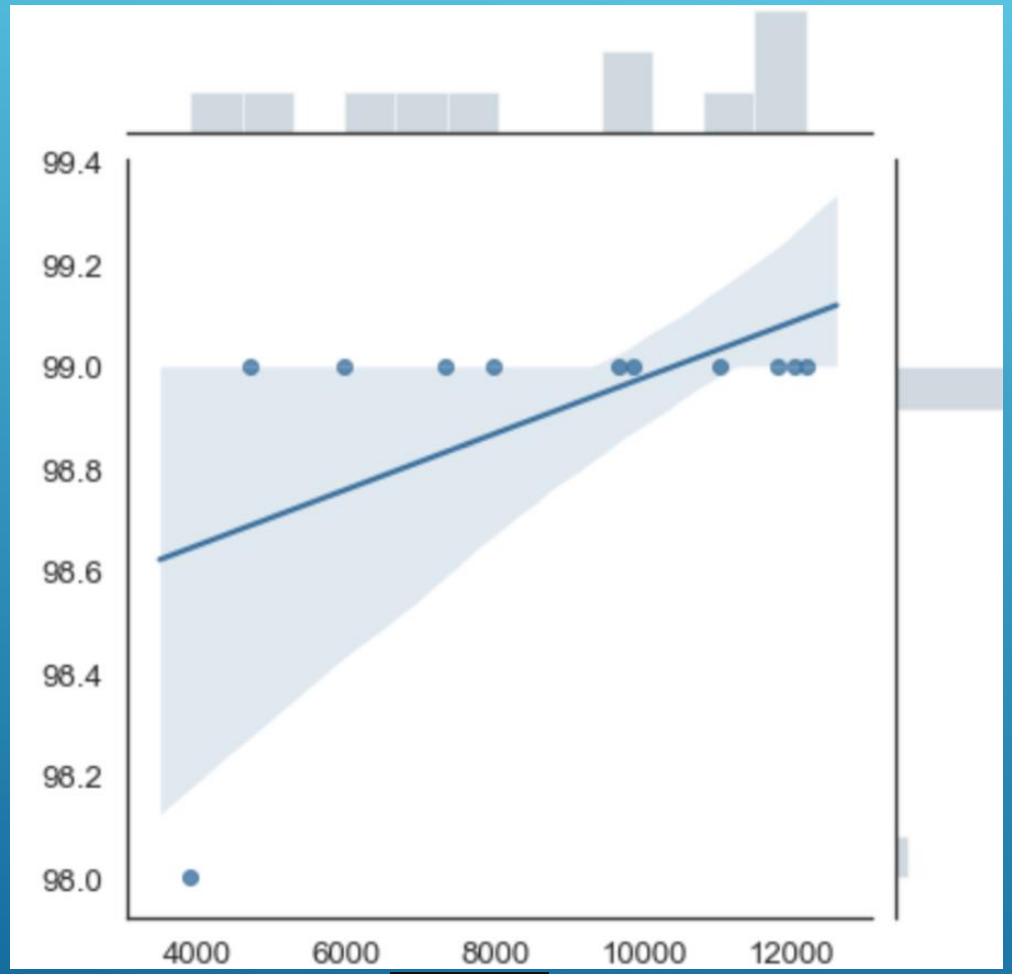


r
0.856

X axis – GNP ($)
Y axis – Measles Immunization Rate

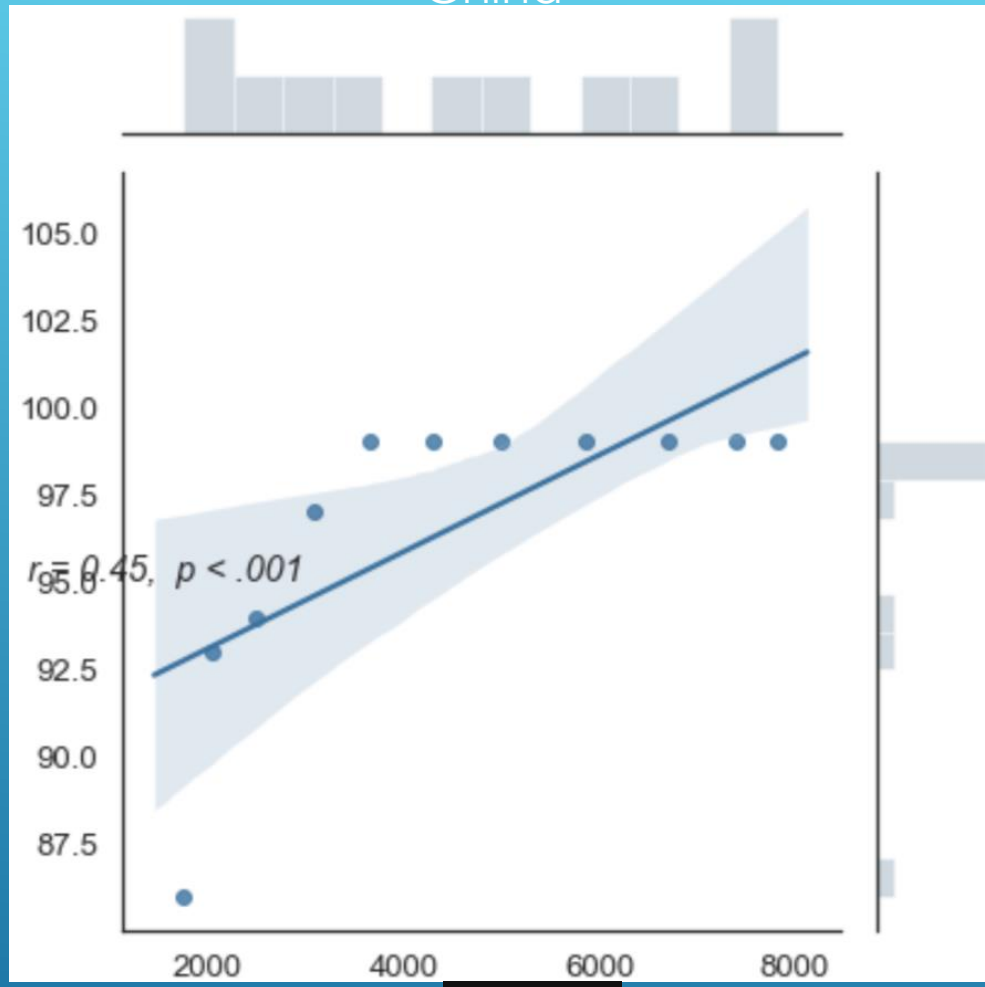
GNP AND IMMUNIZATION RATE CORRELATION

Brazil



r
0.54

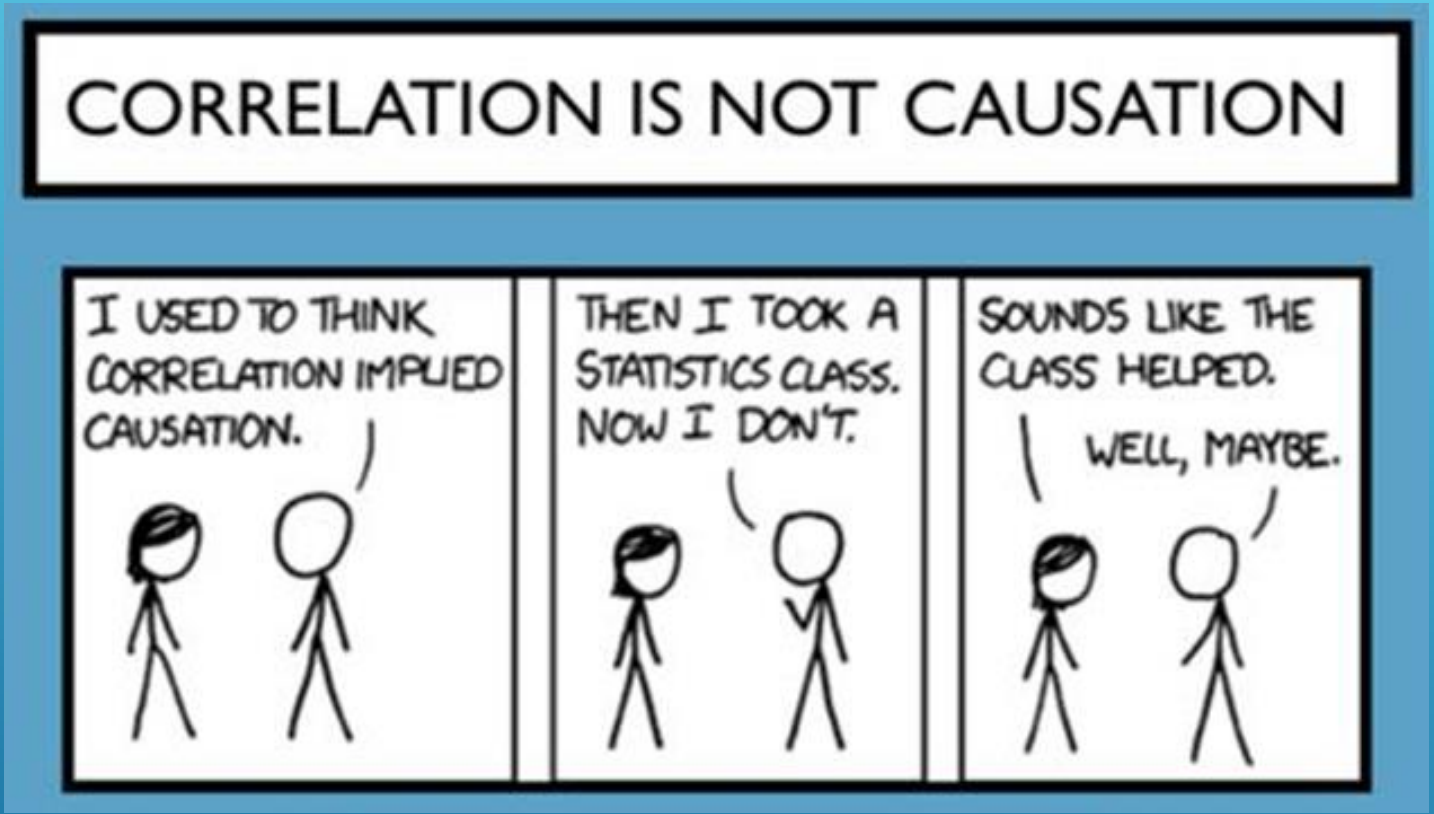
China



r
0.721

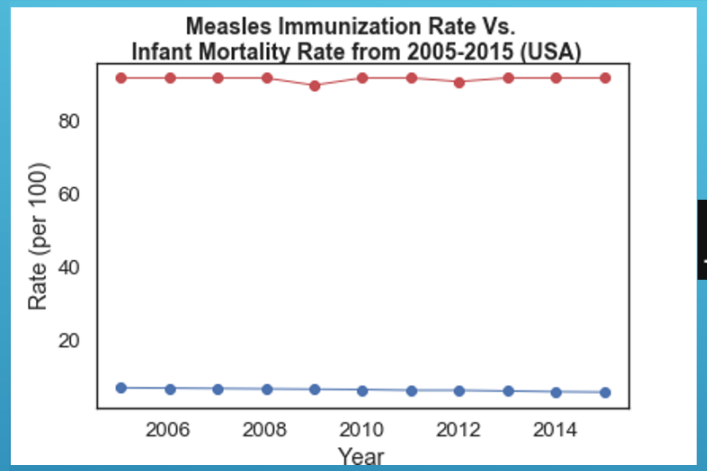
X axis – GNP ($)
Y axis – Measles Immunization Rate

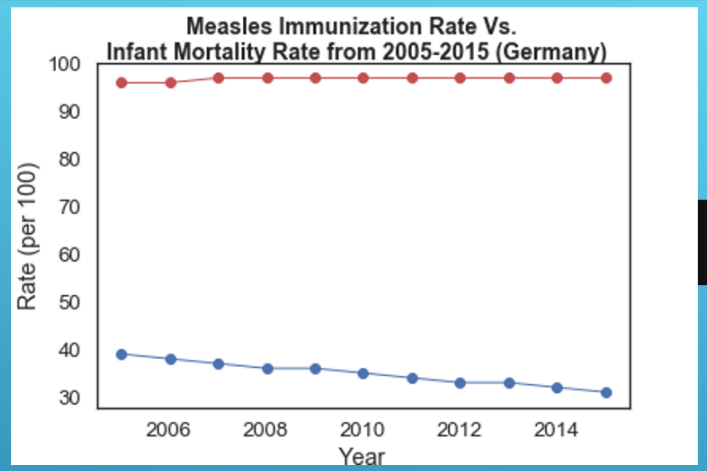
CONCLUSIONS

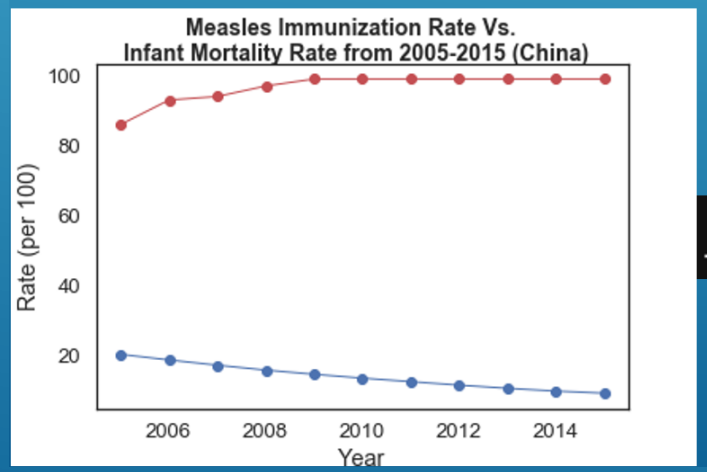


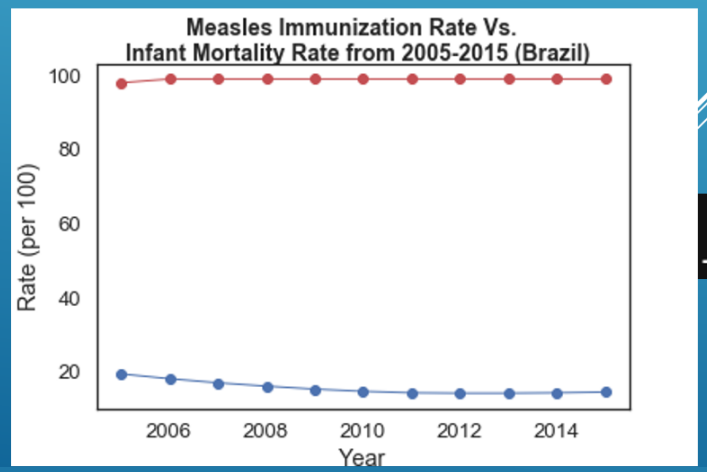
But.....

CONCLUSIONS



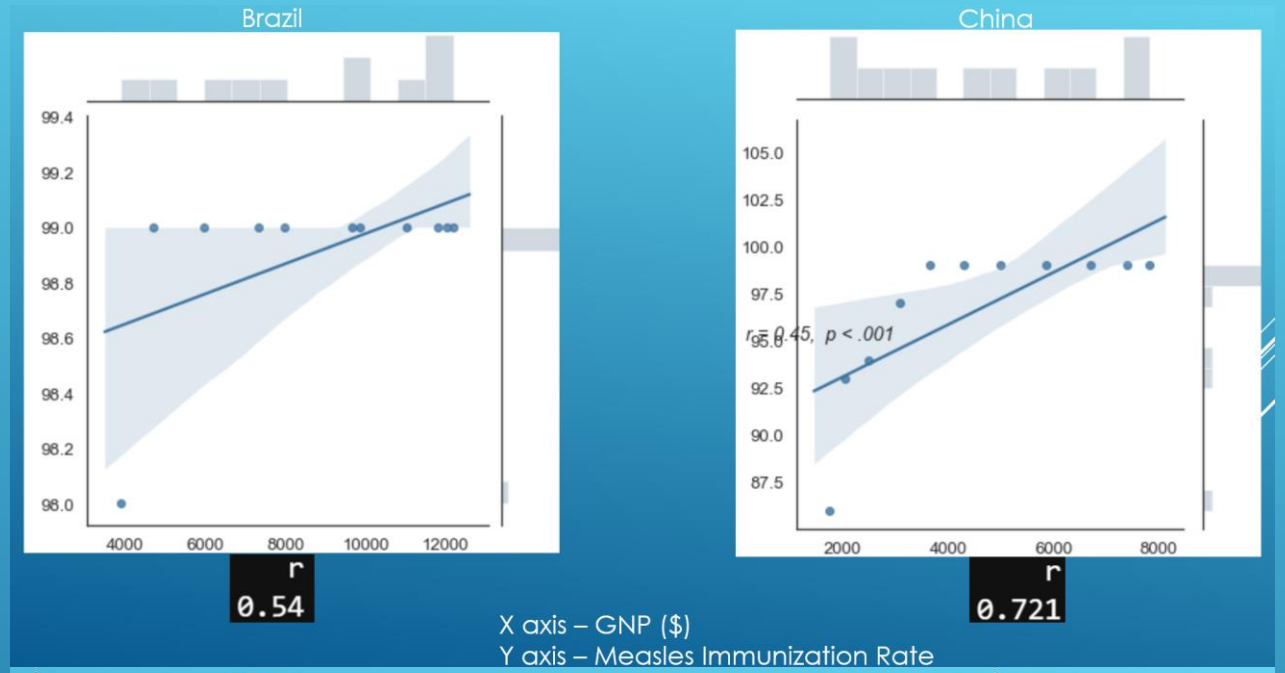






- All four of the countries analyzed had *negative* r values, indicating that while *immunization rates increased, infant mortality rates decreased.*

GNP AND IMMUNIZATION RATE CORRELATION



- All four of the countries analyzed had *positive* r values, indicating that while *GNP rates increased, immunization rates also increased.*

MAKES SENSE – RIGHT?

↑ IMMUNIZATION RATES
↓ INFANT MORTALITY RATES

↑ ECONOMIC ROBUSTNESS (GNP)
↑ IMMUNIZATION RATES

BUT further analysis required to firmly establish causes and factors

QUESTIONS?
COMMENTS?