

WEEKLY COUNTS OF DEATHS

by State and Select Causes, 2014-2018



CENTERS FOR DISEASE
CONTROL AND PREVENTION

National Center for Health Statistics

Columns in the Dataset

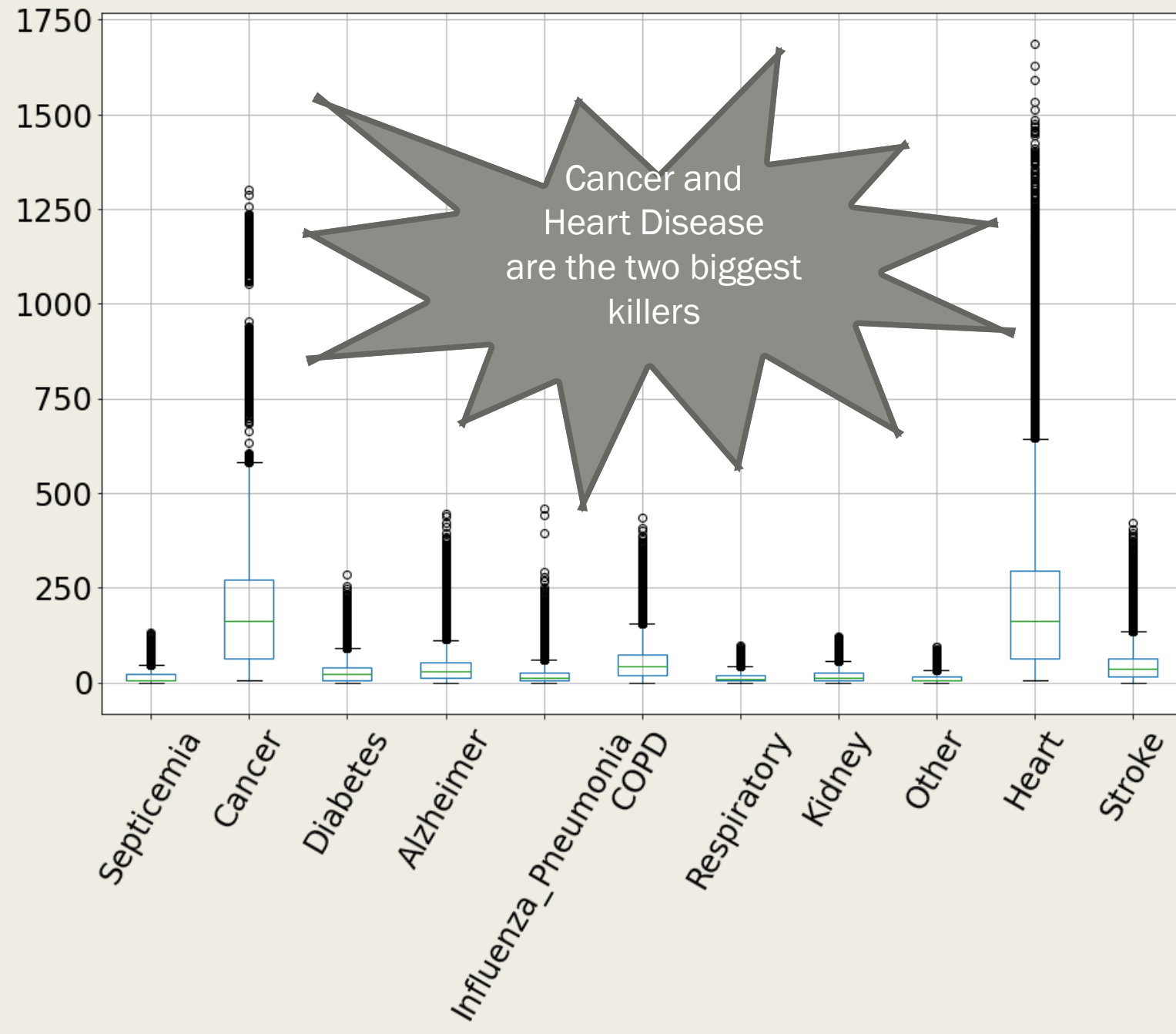
Weekly counts for deaths from:

- Natural Cause
- Septicemia
- Malignant neoplasms (Cancer)
- Diabetes mellitus
- Alzheimer disease
- Influenza and pneumonia
- Chronic lower respiratory diseases
- Other diseases of respiratory system
- Nephritis, nephrotic syndrome and nephrosis
- Other causes
- Diseases of heart
- Cerebrovascular diseases

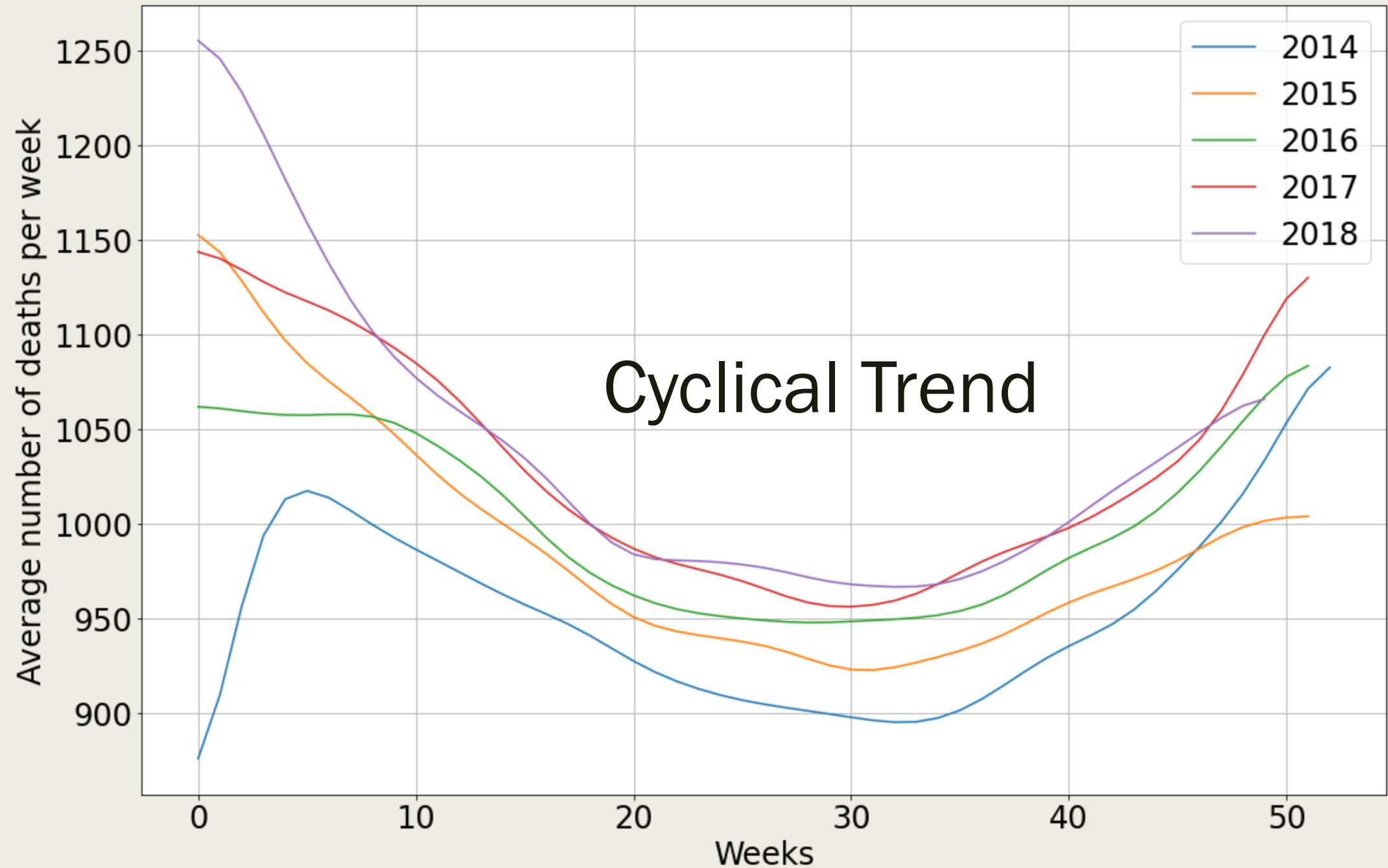
Data broken by

- Jurisdiction (State)
- Week (week ending date)
- Year
- Week Number

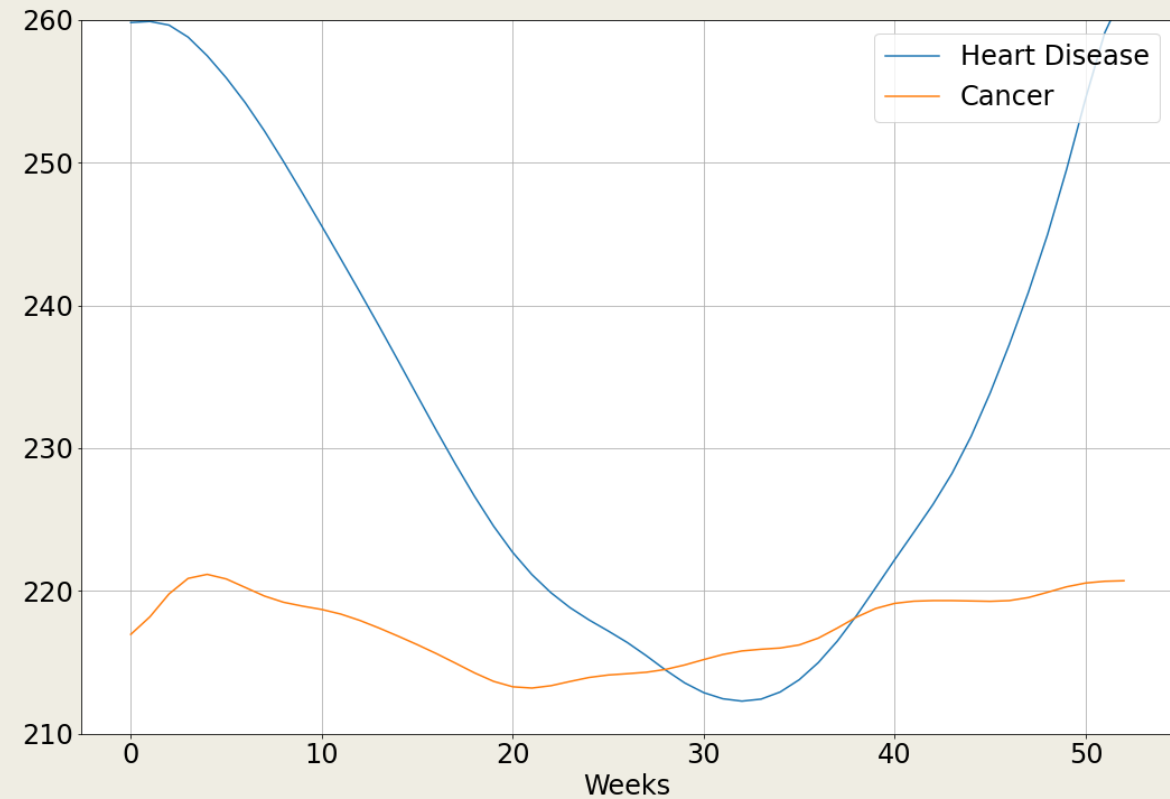
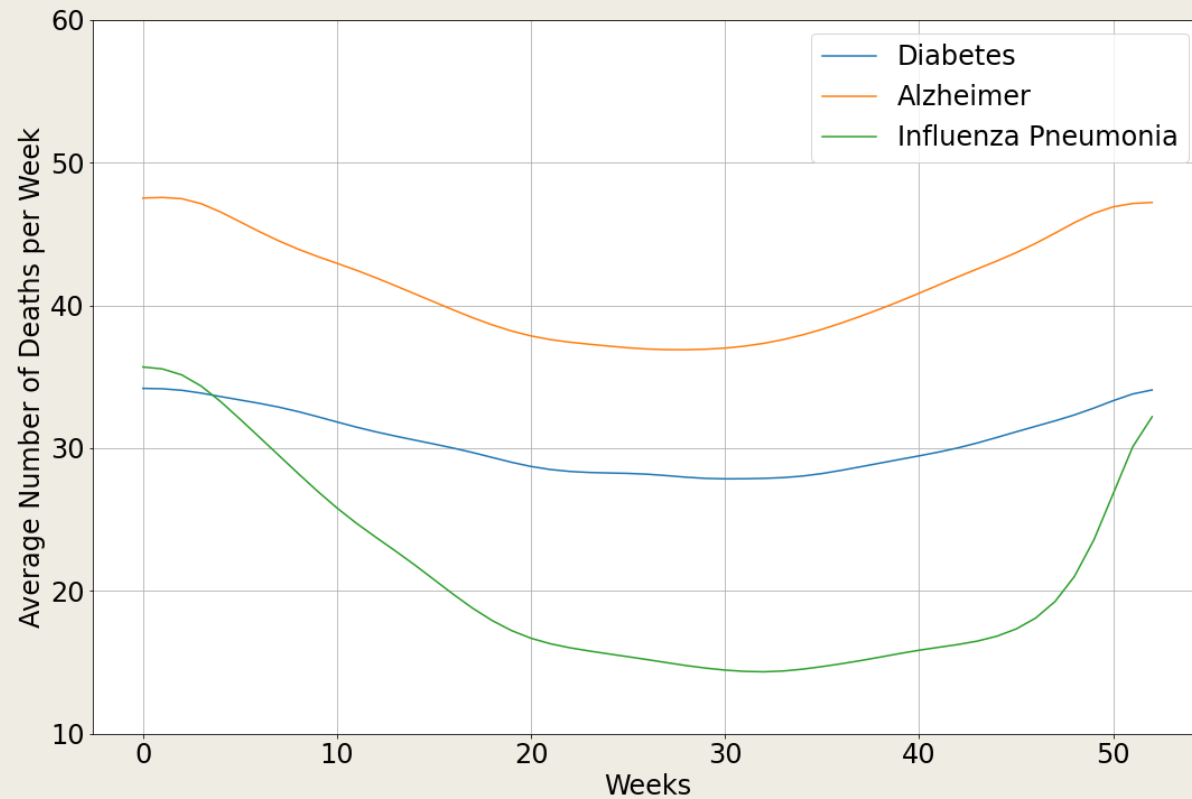
10



All Causes



Cyclical Trend for Various Causes



It appears that various causes of deaths are cyclical as well, although it seems that change in counts for Heart Disease are much higher than Cancer or Diabetes

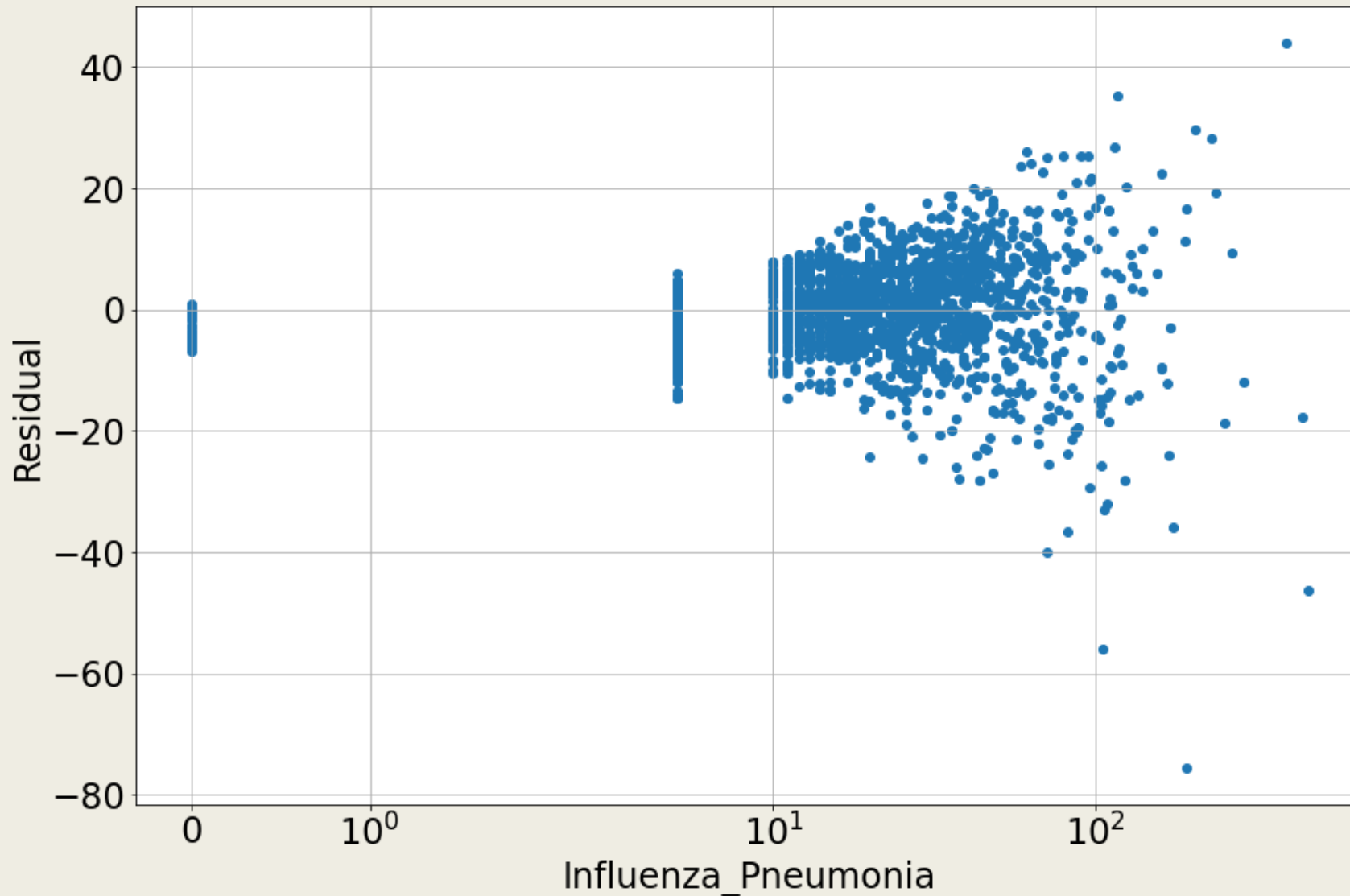
Model Comparison

Model	Features in addition to the base line	Standard Deviation		Root Mean Square Deviation	
		Test	Train	Test	Train
LinearRegression	['Influenza_Pneumonia', 'Respiratory', 'Nat_Cause', 'Septicemia', 'Cancer', 'Diabetes', 'Alzheimer', 'COPD', 'Kidney', 'Other', 'Heart', 'Stroke']	7.06	7.07	7.06	7.07
LinearRegression	['Influenza_Pneumonia', 'Respiratory']	7.32	7.34	7.32	7.33
LinearRegression	['Influenza_Pneumonia', 'Influenza_Pneumonia_SqRoot', 'Respiratory']	7.48	7.24	7.48	7.24
LinearRegression	['Influenza_Pneumonia']	7.56	7.42	7.57	7.42
LinearRegression	['Influenza_Pneumonia', 'Influenza_Pneumonia_Squared', 'Respiratory']	8.60	7.23	8.61	7.23
LinearRegression	Polinomial Features - too many to list	10.86	6.38	11.00	6.38
RandomForestClassifier	['Influenza_Pneumonia', 'Respiratory', 'Nat_Cause', 'Septicemia', 'Cancer', 'Diabetes', 'Alzheimer', 'COPD', 'Kidney', 'Other', 'Heart', 'Stroke']	11.01	4.56	11.32	4.71
LogisticRegression	['Influenza_Pneumonia', 'Respiratory', 'Nat_Cause', 'Septicemia', 'Cancer', 'Diabetes', 'Alzheimer', 'COPD', 'Kidney', 'Other', 'Heart', 'Stroke']	17.61	10.37	18.52	11.73
SGDClassifier	['Influenza_Pneumonia', 'Respiratory', 'Nat_Cause', 'Septicemia', 'Cancer', 'Diabetes', 'Alzheimer', 'COPD', 'Kidney', 'Other', 'Heart', 'Stroke']	21.55	18.20	22.71	19.19
LinearRegression	Base Line	26.96	22.61	27.02	22.61

Base Line: ['Week', 'Year', 'Fall', 'Winter', 'Spring', 'Summer']

Residual Difference Predicted vs Actual

The most accurate model – Linear Regression with the most features



Average number of deaths per week Predicted vs. Actual

The most accurate model – Linear Regression with the most features

