Fatality and
Disease
Analyzation and
Statistics

About the clients

CDC

- The Centers for Disease Control and Prevention is the nation's health protection agency
- They conduct science and provides health information that protects our nation against health threats
- The CDC saves lives

Elder Research

- A data science and predictive analytics company
- Develops analytic solutions and provides analytics consulting

Client objective

 Use US mortality data to find trends in fatal diseases and injuries that co-occur (meaning two or more conditions that are commonly listed together on death certificates as factors that contributed to or caused death).

Understanding the data



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Centers for Disease Control and Prevention National Center for Health Statistics



Instructions for Completing the Cause-of-Death Section of the Death Certificate for Injury and Poisoning (usually completed by a Medical Examiner or Coroner)

Examples of properly completed medical certifications

CAUSE OF DEATH (See instructions and examples)							
32. PART I. Enter the chain of events—diseases, injuries, or complications—that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.						Approximate interval: Onset to death	
IMMEDIATE CAUSE (Final disease or condition	a. Carl	on moi	Unknown				
resulting in death)		Due to (or as a consequence of):					
Sequentially list conditions,	_{b.} Inha	Inhalation of automobile exhaust fumes					
if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE		Due to (or as a consequence of):					
(disease or injury that	C	Due to (or as a consequence of):					
initiated the events resulting in death) LAST	d						
PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I. 33. WAS AN AUTOPSY PERFOR						RMED?	
■ Yes □] No	
					34. WERE AUTOPSY FINDINGS COMPLETE THE CAUSE OF DE		
35. DID TOBACCO USE CONTRIBUTE	TO	36. IF FEM.	ALE:	37. MA	37. MANNER OF DEATH		
		□ Not pregnant within past year					
5 Yes			nant at time of death		☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation		
			regnant, but pregnant within 42 days of death		☐ Accident ☐ Pending Investigation ☐ Suicide ☐ Could not be determined		
E No E CHRISTIII		☐ Unknown if pregnant within the past year				Clermined	
38. DATE OF INJURY	39. TIME O	IE OF INJURY 40. PLACE OF INJURY (e.g., Decedent's home; construction site; restaurant; wooded area) 41.				1. INJURY AT WORK?	
(Mo/Day/Yr) (Spell Month) May 5, 2003	Unknov	wn	Own home garage			Yes No	
42. LOCATION OF INJURY:	State: Mis	souri	City or Town: near Alexandria				
Street & Number: 898 Sylvan Road Apartment No.: Zip Code: 65100-1234							
43. DESCRIBE HOW INJURY OCCURRED: 44. IF TRANSPORTATION ACCII						ACCIDENT, SPECIFY:	
Inhaled carbon monoxide from auto exhaust through hose in an enclosed garage							
□ Pedestrian							
					☐ Other (Specify)		

Understanding the data (cont.)

AC	AD	AE	AF	AG	AH	Al	AJ	AK
econdp_1	econds_1	enicon_1	econdp_2	econds_2	enicon_2	econdp_3	econds_3	enicon_3
:	1 1	1500	6	1	L031			
1	1 1	1469	2	1	R042	3	1	C349
	1 1	G309						
	1 1	T71	1	2	X91	2	1	T71
	1 1	1250	2	1	S720	6	1	X590
1	1 1	1499	2	1	1516	6	1	E780
1	1 1	E274						
	1 1	1500						
	1 1	1500	2	1	1350			
	1 1	T142	2	1	W19	3	1	R688
1	1 1	J189	2	1	148	6	1	164
	1 1	G309	6	1	1500			
	1 1	1219	2	1	1251	6	1	N19
	1 1	R688	1	2	R54			
:	1 1	C798	6	1	K769	6	2	F179
:	1 1	S099	1	2	X599			
	1 1	T71	1	2	X70	2	1	T71
	1 1	S069	2	1	S019	2	2	X72
	1 1	164	2	1	R13			
	1 1	J189	6	1	G20	6	2	164
	1 1	1250	6	1	J449			
	1 1	1500	2	1	E668			
	1 1	A419	2	1	J449	6	1	F179
1	1 1	1469	2	1	R64	2	2	E86

ICD 10

10th revision of the <u>International</u>
Statistical Classification of Diseases and
Related Health Problems (ICD), a <u>medical</u>
classification list by the <u>World Health</u>
Organization(WHO).

It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases.

The code set allows more than **14,400 different codes** and permits the tracking of many new diagnoses.

Understanding the data (cont.)

- Used mortality data of 10 years (2006-2015)
- Approx 25 million records
- Approx 10 GB data
- A record can contain up to 20 causes of death (specified as 4-char 'ICD' codes)
- To find patterns of co-occurring codes, 'association rules' to be used

Technical setup

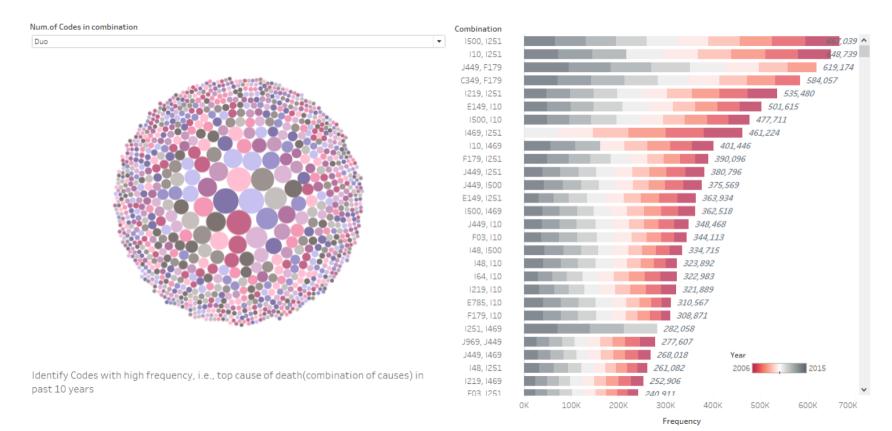
- R for data cleaning
- Python with Spark on AWS for scalable analysis (as data increases YoY)
- 1 master, 10 worker nodes setup on AWS (each 16CPU, 30GB memory)
- FPGrowth (Frequent Pattern Growth) algorithm
 - o 'Support' parameter (used 0.0001)
 - 'Partitions' parameter (used 480)
- Tableau for visualizations

Key Findings

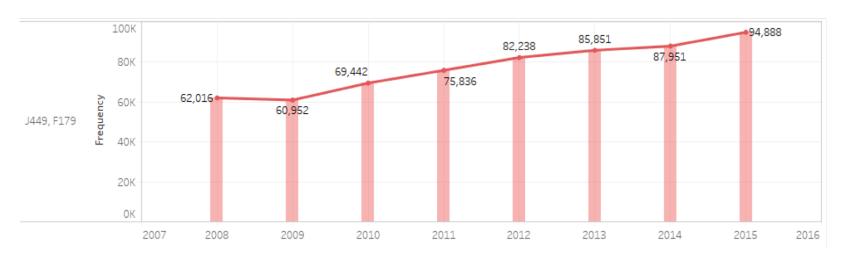
Top co-occurring code pairs

Rank	Combina	ation	Code1 Desc	Code2 Desc	Frequency
1	1500	I251	Congestive heart disease	Atherosclerotic heart disease, Coronary	667,039
2	I10	I251	Essential (primary) hypertension	Atherosclerotic heart disease, Coronary	648,739
3	J449	F179	Chronic obstructive pulmonary disease	Mental/behavioural disorders due to use of tobacco	619,174
4	C349	F179	Malignant neoplasm: Bronchus or lung	Mental/behavioural disorders due to use of tobacco	584,057
5	I219	I251	Acute myocardial infarction	Atherosclerotic heart disease, Coronary	535,480
6	E149	I10	Diabetes mellitus without complications	Essential (primary) hypertension	501,615
7	1500	I10	Congestive heart disease	Essential (primary) hypertension	477,711
8	I469	I251	Cardiac arrest	Atherosclerotic heart disease, Coronary	461,224
9	I10	1469	Essential (primary) hypertension	Cardiac arrest	401,446
10	F179	I251	Mental/behavioural disorders due to use of tobacco	Atherosclerotic heart disease, Coronary	390,096

Top co-occurring code pairs



Trends in (selected) co-occurring code pairs

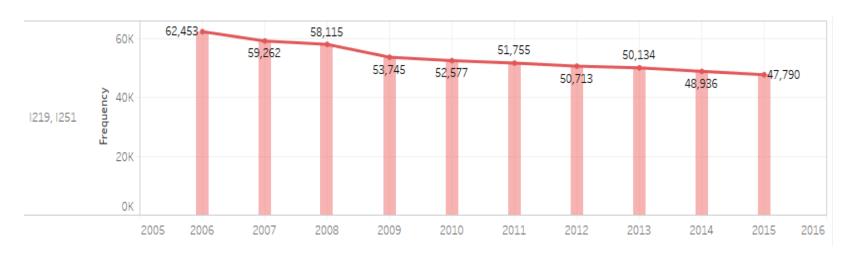


Highest increase (~62k to ~95k):

J449 - Chronic obstructive pulmonary disease, unspecified

F179 - Mental and behavioural disorders due to use of tobacco: Unspecified mental and behavioural disorder

Trends in (selected) co-occurring code pairs

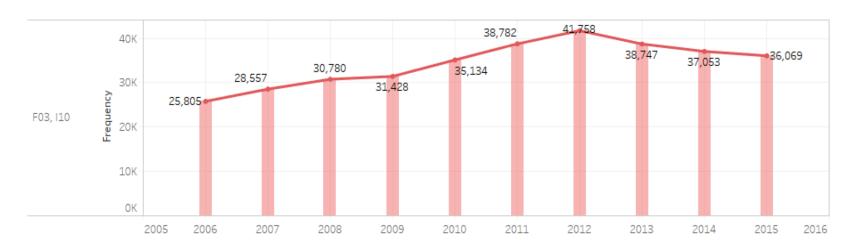


Highest decrease (~62k to ~48k):

1219 - <u>Acute Myocardial Infarction, Unspecified</u>

1251 - Atherosclerotic Heart Disease

Trends in (selected) co-occurring code pairs



Most interesting (increasing till 2012 and then decreasing):

F03 - Unspecified dementia

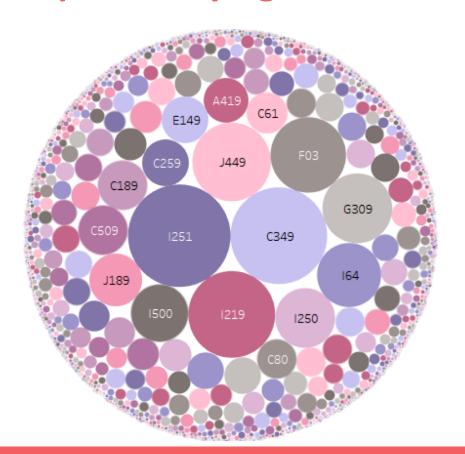
110 - Essential (primary) hypertension

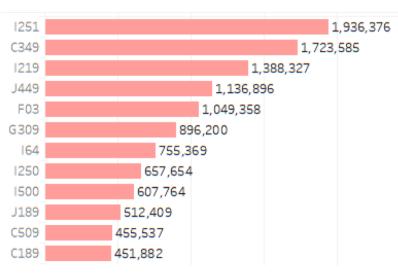
Other Findings

Top 'underlying causes of death'

Rank	UCOD	Description	Count
1	I251	Atherosclerotic heart disease of native coronary artery	1,936,376
2	C349	Malignant neoplasm: Bronchus or lung, unspecified	1,723,585
3	I219	Acute myocardial infarction, unspecified	1,388,327
4	J449	Chronic obstructive pulmonary disease, unspecified	1,136,896
5	F03	Unspecified dementia	1,049,358
6	G309	Alzheimer's disease, unspecified	896,200
7	164	Stroke, not specified as haemorrhage or infarction	755,369
8	1250	Atherosclerotic cardiovascular disease	657,654
9	1500	Congestive heart failure	607,764
10	J189	Pneumonia	512,409

Top 'underlying causes of death'





Trends in (selected) codes



Top occurring codes where injury was involved

X44: Accidental poisoning by and exposure to other and unspecified drugs, medicaments and biological substances 245,222	W19: Unspecified fall 195,526	X74: Intentional self-harm by other and unspecified firearm discharge 94,700		
X42: Accidental poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified 227,932	X95: Assault by other and unspecified firearm discharge - 167,212	X70: Intentional self-harm by hanging, strangulation and suffocation 75,929	W80:	
	W18: Other fall on same level 128,058	F03: Unspecified dementia		
		X590: Exposure to unspecified factor		

Thank You