
INST452:

Health Data Analytics

Module 3
Health Data Sources & Analytics

Dr. Nikki Sigalo

Health Data Sources

- Surveys
- Medical records
- Claims data
- Vital records
- Surveillance data/disease registries

Surveys

- Methods
 - Self-administered questionnaires
 - In-depth interviews via phone and in-person
- Types
 - Population
 - Behavioral Risk Factor Surveillance System (BRFSS)
 - National Health and Nutrition Examination Survey (NHANES)
 - Provider
 - National Ambulatory Medical Care Survey



Section 6: Chronic Health Conditions

Has a doctor, nurse, or other health professional EVER told you that you had any of the following? For each, tell me "Yes," "No," or you're "Not sure."

6.1 (Ever told) you that you had a heart attack also called a myocardial infarction? (106)

1	Yes
2	No
7	Don't know / Not sure
9	Refused

6.2 (Ever told) you had angina or coronary heart disease? (107)

1	Yes
2	No
7	Don't know / Not sure
9	Refused

6.3 (Ever told) you had a stroke? (108)

1	Yes
2	No
7	Don't know / Not sure
9	Refused

6.4 (Ever told) you had asthma? (109)

1	Yes
2	No
7	Don't know / Not sure
9	Refused

[GO TO Q6.6]

6.5 Do you still have asthma? (110)

1	Yes
2	No
7	Don't know / Not sure
9	Refused

[GO TO Q6.6]

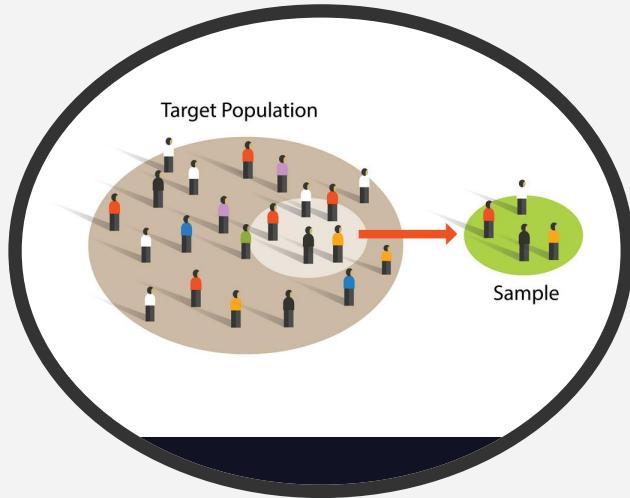
Surveys in Use – Data Analytics

Table 5 Medical care access among transgender and non-transgender respondents, BRFSS, 2016

Medical care access	Yes, male to female (%)	Yes, female to male (%)	Yes, non-conforming (%)	Non-transgender (%)
Difficulty to see doctor due to cost in the past 12 months (P=0.000)				
Yes	50 (15.7)	41 (17.3)	24 (16.3)	18,085 (9.2)
No	269 (84.3)	196 (82.7)	123 (83.7)	178,898 (90.8)
Total	319 (100.0)	237 (100.0)	147 (100.0)	196,983 (100.0)
Visited a doctor for a routine checkup (P=0.505)				
Within past year	242 (76.1)	172 (75.1)	105 (71.9)	150,057 (76.8)
More than 1 year ago	76 (23.9)	57 (24.9)	41 (28.1)	45,436 (23.2)
Total	318 (100.0)	229 (100.0)	146 (100.0)	195,493 (100.0)
Had flu shot or vaccine in 12 months (P=0.087)				
Yes	129 (40.2)	96 (41.4)	60 (41.4)	89,323 (45.6)
No	192 (59.8)	136 (58.6)	85 (58.6)	106,393 (54.4)
Total	321 (100.0)	232 (100.0)	145 (100.0)	195,716 (100.0)

BAFSS, Behavioral Risk Factor Surveillance System.

Pros



Collect health and social science information from a sample of people in a standardized way to better understand a larger population

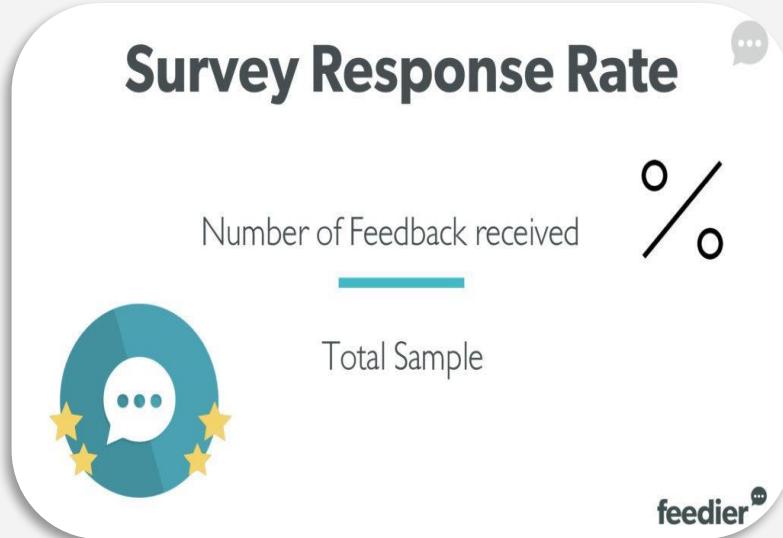


Collect empirical data in a relatively short period of time

Cons



Can be hard to get detailed information in a survey (limited to static survey responses)



Surveys can have low response rates

Cons



Sometimes people choose not to answer difficult questions or they cannot remember important details correctly or at all (recall bias)

Those who do not have access to the medium through which the surveys are distributed are excluded



Medical Records/Electronic Health Records (EHRs)

- Track events and transactions between patients and health care providers
- Offer information on diagnoses, procedures, lab tests, and other services
- Help us measure and analyze trends in health care use, patient characteristics, and quality of care

BURNA BOY (DOB: 7/2/1991) 32 YEAR OLD MALE

File Edit View Help Know

Demographics | Summary Sheet | Most Recent Encounter | Past Encounters | Imported Items | Account Information

Sunday April 24, 2011 09:35 AM

Use a prior visit as a stepping off point for this visit. Overwrite all fields Keep existing cc, hpi, & ros

Chief Complaint: Annual Exam: History of Present Illness: Feeling fine, though hasn't been taking his bp medication as his legs have been swelling up.

Review of Systems: Constitutional: Negative Eyes: Negative Ears, Nose, Throat: Negative

Past Medical History: h/o Rheumatic Fever h/o HTN

Social History: Former smoker Pks/Yrs 20 Yrs Quit 10 Some alcohol use with Neg. CAGE. Works for the NY Times - stressful recently as cutting back due to economy. Tobacco: Former smoker (20 pk yrs / 10 yrs quit)

Family History: Patient denies any family history of CAD, HTN, DM, or CA.

Allergies: NKDA (Updated by JONATHAN on 04/24/2011 09:39 AM)

Current Medications: AMLODIPINE 5mg, 1 po qd PROZAC 20 MG, One tablet Daily

Physical Exam: Note: 190 lb 74 in Date: 03/04/2009 Weight lb: 208.0 Height in: 62.0 Temp F: 98.7 F BP: 180/85 Pulse: 92 RR: 16 BMI: 38.1 Date: 04/06/2008 Height in: 62.0 Date: 11/02/2003 Weight lb: 205.0 Height in: 62.0 BMI: 37.6

Pictures: Illustrations: Confidential:

GENERAL: WNWD NAD HEENT: WNL LUNGS: CTA HEART: RRR S1 S2 without murmurs, thrills, rubs CHEST WALL: WNL

Diagnoses: Search Dx (ICD9): Problem List: No Active Diagnoses Knowledge Base

Assessment: # ROUTINE GENERAL MEDICAL EXAMINATION (V70.0): Reviewed common prevention issues. Had flu shot previously. Wears seatbelts. # HYPERTENSION NOS (401.9): Advised restart medication. He doesn't want to change med now despite swelling.

Plan: Decision Support: Patient Education Given: Name: Screening for Lipid Disorders in Adults (Male) Next: 12/01/1980 Screening for Obesity in Adults Next: 12/01/1963 Td (Tetanus, diphtheria) After 12/01/19 Typhoid 06/01/1946

Forward Chart: Sign-Off: Right-click on a recommendation for more information or to edit the rule.

65 year old man last seen 2 years ago (03/04/09) by Jonathan Bertman, MD

JONATHAN 4/24/2011

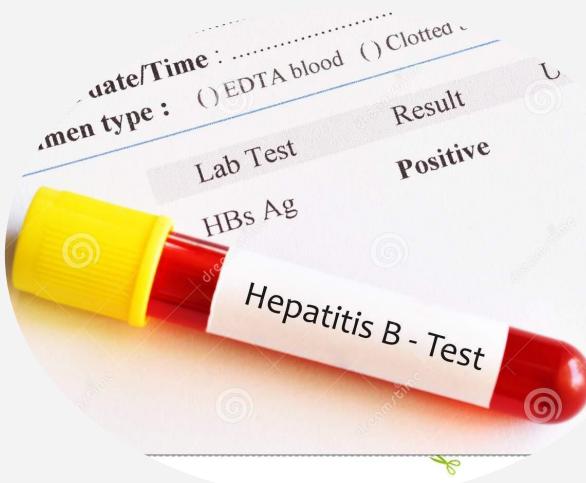
EHR in Use – Data Analytics

Age Diagnosis	All ages	0-17	0-4	5-17	5-9	10-17
All Diagnoses	1,165.3	337.6	585.1	238.6	167.2	283.8
Infectious and parasitic diseases	42.7	16.7	*	7.3	8.3	6.7
Septicemia	24.5	1.8	4.4	*	*	*
Cancer, all	41.6	*	*	*	*	*
Endocrine, nutritional and metabolic disorders	59.3	18.9	*	12.4	11.1	13.2
-Diabetes	21.2	3.9	*	5.0	*	6.3
-Dehydration	10.9	9.1	23.3	3.3	4.7	2.5
Anemias	13.4	4.0	4.2	4.0	2.6	4.9
Mental disorders	68.2	32.0	*	44.0	12.6	63.8
-Alcohol and drug	14.4	1.5	*	2.1	*	3.4
Nervous system disorders	33.0	15.8<	28.0	*	10.8	*
Circulatory system disorders	197.3	*	*	*	*	*
-Heart disease	127.9	*	*	*	*	*
--Arrhythmias	26.1	0.6	*	*	*	*
--Heart failure	33.8	*	*	*	*	*

Pros



Medical records are usually accurate and detailed because they come from straight health care providers



Can be used for research, like comparing how effective providers are, and seeing how patients respond to treatment



The data include information that patients might not think to add or feel comfortable sharing through other data sources like surveys

Cons



Medical records are (by definition) only available for people who are able to get medical care

Claims Data

- Claims databases collect information on millions of doctors' appointments, bills, insurance information, and other patient-provider communications

NDE	# of CCs	# of Enrollees Part A	# of Enrollees Part B	Average Payment Part A	Average Payment Part B
0	10,138,926	7,497,739	\$248	\$1,156	
1	6,663,517	6,498,765	\$1,314	\$2,722	
2	4,583,587	4,514,823	\$2,998	\$4,258	
3	2,632,736	2,606,318	\$5,968	\$6,128	
4	1,399,364	1,389,361	\$10,784	\$8,472	
5	649,251	646,544	\$17,537	\$10,974	
6	251,404	250,820	\$26,153	\$13,597	
7	80,674	80,613	\$36,243	\$16,283	
8	19,532	19,543	\$46,766	\$18,729	
9	2,991	2,996	\$56,014	\$20,638	
10	225	225	\$68,333	\$22,367	

Pros



Come directly from notes made by the health care provider, and the information is recorded at the time patient sees the doctor

Because of the large sample size of claims data, researchers can analyze groups of patients with rare illnesses and medical conditions



Cons



There may be low validity due to certain illegal billing practices, like ordering unnecessary tests or billing for services that were not provided

Knowledge Check

- ▶ Navigate to Socrative.com
- ▶ Click “Login” in the top right-hand corner
- ▶ Click “Student Login”
- ▶ Enter INST452 for the Room Name and click “JOIN”
- ▶ Enter your full name (LastName, FirstName) and click “DONE”
- ▶ i.e. Sigalo, Nekabari

Vital Records

- Collected by the National Vital Statistics System, and are maintained by state and local governments
- Include births, deaths, marriages, divorces, and fetal deaths
- Record information about the cause of death, or details of the birth

STATE OF FLORIDA

OFFICE of VITAL STATISTICS

CERTIFIED COPY

CERTIFICATE OF DEATH STATE FILE NO. **'60-023882**

FLORIDA REGISTRAR'S NO. **03-1960**

BIRTH NO. **26-08**

PLACE OF BIRTH **Duval** CODE NO. **26-08** 2. RESIDENCE (Where deceased lived) **Florida** 11. (Indicates residence before admission)

4. CITY, TOWN OR LOCATION **Jacksonville** 5. IS PLACE OF DEATH **Yes** **No** 6. CITY, TOWN, OR LOCATION **Jacksonville** 8. COUNTY **Duval**

4. NAME OF HOSPITAL OR INSTITUTION **Very Hospital** 7. LENGTH OF ILLNESS **10 days** 9. STREET ADDRESS **1438 Harrison St.** 10. IS RESIDENCE **Yes** **No** **Yes** **No** **Yes** **No**

11. FATHER'S NAME **John Smith** 12. AGE (In years) **59** 13. MOTHER'S NAME **William J. Smith** 14. CITIZEN OF WHAT COUNTRY? **USA**

15. COLOR OR RACE **Colored** 16. MARRIED **Yes** **No** **Never Married** 17. DATE OF BIRTH **June 22, 1890** 18. DATE OF DEATH **June 1960**

19. OCCUPATION (Give kind of work done during past year) **Fireman, ret.** 20. EMPLOYER'S NAME **Lumber Co.** 21. PLACE OF BUSINESS OR INDUSTRY **Swainsboro, Georgia USA** 22. PLACE OF DEATH **Swainsboro, Georgia USA**

23. FATHER'S NAME **Solomon Smith** 24. MOTHER'S NAME **William J. Smith**

25. WAS DECEASED EVER IN U.S. ARMED FORCES? **No** 26. SOCIAL SECURITY NO. **264-03-2031** 27. EMPLOYER'S SIGNATURE **William J. Smith**

28. PART OF OTHER OCCUPATION **Retired** 29. ADDRESS **1438 Harrison Street** 30. INTERNAL BETWEEN DEATH AND DEATH

31. IMMEDIATE CAUSE (a) **Heart Disease** (b) **Hypertension** (c) **Arteriosclerosis** 32. DATE OF DEATH **June 1960** 33. DATE OF AUTOPSY **NO**

34. MEDICAL CERTIFICATION

35. DECEASED IF YES **No** 36. DATE DUE TO (a) **10:45 AM** 37. DATE DUE TO (b) **10:45 AM** 38. DESCRIBE HOW DEATH OCCURRED **Deceased from heart attack at home** 39. DATE OF DEATH **June 1960** 40. DATE OF AUTOPSY **NO**

41. DEATH OCCURRED **At home** 42. PLACE OF DEATH **6-Park** 43. CITY, TOWN, OR LOCATION **Jacksonville** 44. COUNTY **Duval** 45. STATE **Florida**

46. DEATH OCCURRED **On the date stated above; and to the best of my knowledge, from the causes stated** 47. DATE OF DEATH **June 11, 1960** 48. DATE OF AUTOPSY **NO**

49. BURIAL (Cremation) **Burial** 50. DATE **6-13-1960** 51. NAME OF CEMETERY OR CEMINATRY **Evergreen Cemetery** 52. LOCATION (City, State, or County) **Jacksonville, Florida**

53. BURIAL (Cremation) **Cremated** 54. DATE **June 11, 1960** 55. DATE NEEDED BY LOCAL REC. **June 11, 1960** 56. RECORDED SIGNATURE **William J. Smith**

57. BURIAL (Cremation) **Holmes Funeral Directors** 58. DATE **June 11, 1960** 59. DATE NEEDED BY LOCAL REC. **June 11, 1960** 60. RECORDED SIGNATURE **William J. Smith**

State Registrar Date Issued: MAR 19 2008

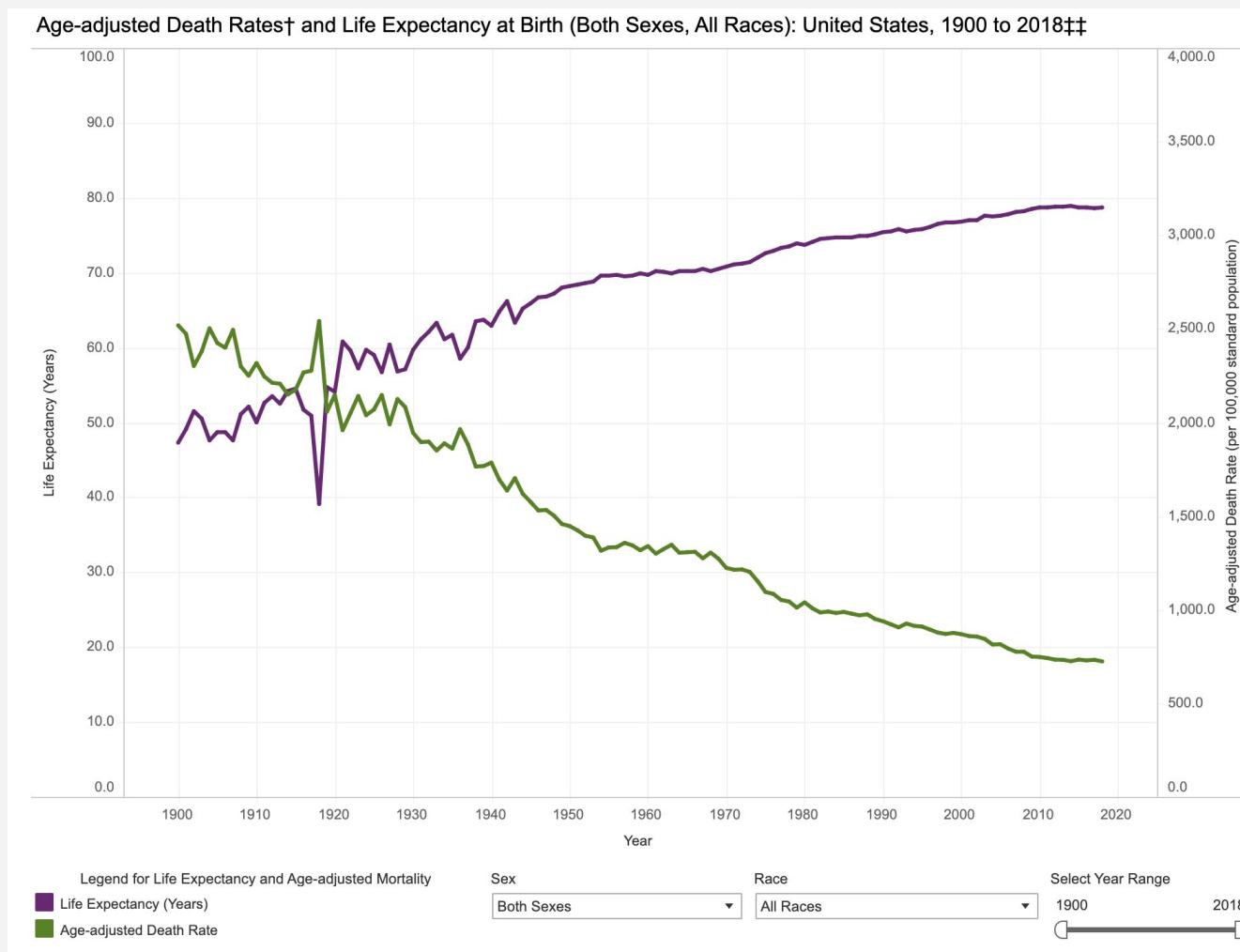
THE ABOVE SIGNATURE CERTIFIES THAT THIS IS A TRUE AND CORRECT COPY OF THE OFFICIAL RECORD ON FILE IN THIS OFFICE.
WARNING: THIS DOCUMENT IS PRINTED OR PHOTOCOPIED ON SECURITY PAPER WITH A WATERMARK OF THE GREAT SEAL OF THE STATE OF FLORIDA. IT IS AN OFFENSE TO FRAUDULently USE, ALTER, OR COPIE THIS DOCUMENT. IT IS ALSO AN OFFENSE TO FRAUDULently USE, ALTER, OR COPIE THE GREAT SEAL OF THE STATE OF FLORIDA.
THE GREAT SEAL OF THE STATE OF FLORIDA
IN GOD WE TRUST

FLORIDA DEPARTMENT OF
HEALTH

OH FORM 1962 (09-04)

24605768 CERTIFICATION OF VITAL RECORD

Vital Records in Use – Data Analytics

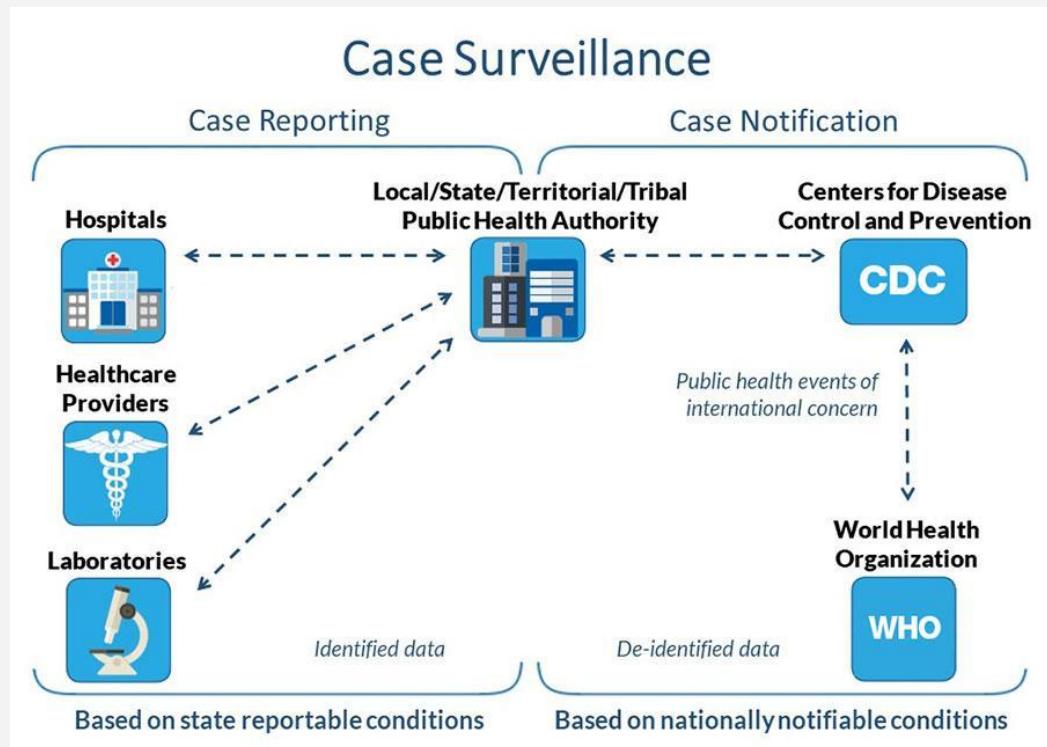


Vital Records

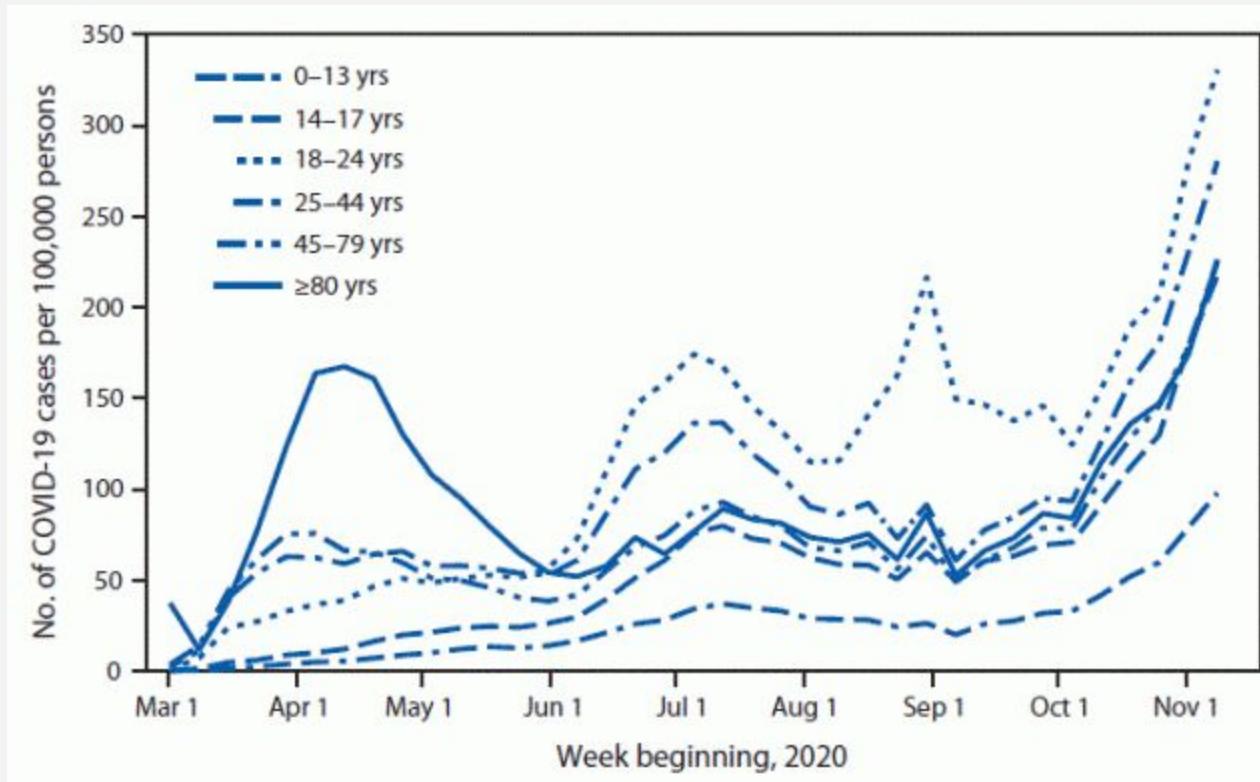
- Pros
 - Offer very detailed information and include information about rare disorders **that end in death**
- Cons
 - Unfortunately, because there are so many state and local governments collecting this information, records can be inconsistent
 - Vital records only provide information on diseases and illnesses that end in death

Surveillance Data

- Public health surveillance is the ongoing systematic collection, analysis, and interpretation of data
- Usually associated with the study of infectious diseases
- The CDC, WHO, and many other institutions operate databases and automated electronic reporting systems to track and monitor outbreaks of specific diseases, like COVID-19 and HIV



Surveillance Data in Use – Data Analytics



Surveillance Data

- Pros
 - Surveillance data has a higher validity than surveys, because the data comes from lab tests, diagnoses, and other patient records
 - Registries also make this data easy to store and analyze
- Cons
 - Because diseases sometimes change definitions, it can be difficult to accurately track trends
 - Data can also be lacking if hospitals or doctors do not report it

Documentation/Data Dictionary

- Every data source should come with some documentation
- Data dictionary/codebook
 - Provides information on the structure, contents, and layout of a **data** file
- Example
 - https://www.cdc.gov/brfss/annual_data/annual_2021.html
 - <https://www.cdc.gov/brfss/questionnaires/pdf-ques/2019-BRFSS-Questionnaire-508.pdf>

A	B	C	D	E
Patient_ID	Patient_Name	Age	BPHIGH4	BPMEDS
1111	Jane Doe	34	2	9
2222	John Smith	54	7	9
3333	Ashley Williams	22	9	2
4444	Daniel Wiggins	36	1	1
5555	Tia Lowe	65	7	1
7777	Angel Hughes	33	4	7

Core Section 4: Hypertension Awareness

Question Number	Question text	Variable names	Responses (DO NOT READ UNLESS OTHERWISE NOTED)	SKIP INFO/ CATI Note	Interviewer Note (s)	Column(s)
C04.01	Have you ever been told by a doctor, nurse, or other health professional that you have high blood pressure?	BPHIGH4	1 Yes 2 Yes, but female told only during pregnancy 3 No 4 Told borderline high or pre-hypertensive 7 Don't know / Not sure 9 Refused	Go to next section	If "Yes" and respondent is female, ask: "Was this only when you were pregnant?" By other health professional we mean nurse practitioner, a physician assistant, or some other licensed health professional.	112
C04.02	Are you currently taking prescription medicine for your high blood pressure?	BPMEDS	1 Yes 2 No 7 Don't know / Not sure 9 Refused			113

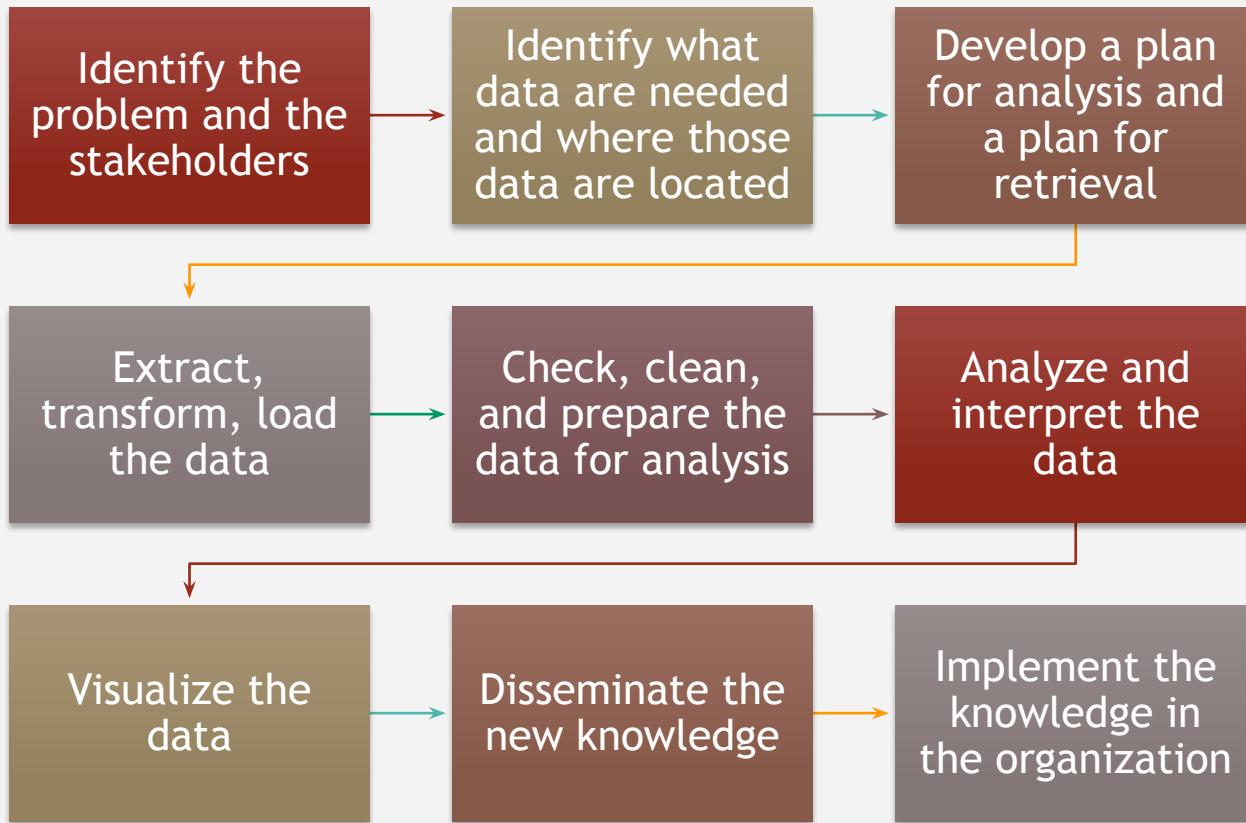
Knowledge Check

- ▶ Navigate to Socrative.com
- ▶ Click “Login” in the top right-hand corner
- ▶ Click “Student Login”
- ▶ Enter INST452 for the Room Name and click “JOIN”
- ▶ Enter your full name (LastName, FirstName) and click “DONE”
- ▶ i.e. Sigalo, Nekabari



10-minute break

Steps in Data Analytics



1. Identify the Research Question and the Stakeholders



Why is this an important problem?



How will the results impact patient care or the institution?



What is the business case?

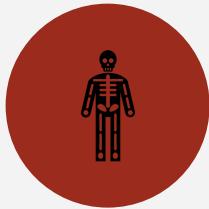


Who are the stakeholders?

Let's Choose a Research Question

- ▶ Navigate to Socrative.com
- ▶ Click “Login” in the top right-hand corner
- ▶ Click “Student Login”
- ▶ Enter INST452 for the Room Name and click “JOIN”
- ▶ Enter your full name (LastName, FirstName) and click “DONE”
- ▶ i.e. Sigalo, Nekabari

2. Identify What Data Are Needed



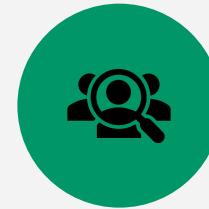
What data elements, such as date of birth, gender, medications, laboratory results, and so on, are needed?



Where are these data elements located—in which system or systems and which database tables?



Is there a clinical data warehouse?



Who is the contact person for each system who will be responsible for retrieving the data?

3. Develop Plans for Retrieval and Analysis

- **Retrieval**
 - Enlist database administrator for each system
 - Develop specific plan for retrieving the required data elements
 - Method for cross-checking number of records as well as completeness—how many should you expect and did you get everything?
- **Analysis**
 - Enlist statistician
 - Identify population, sample size, statistical tests to be performed



4. Extract, Transform, Load Process

Extraction

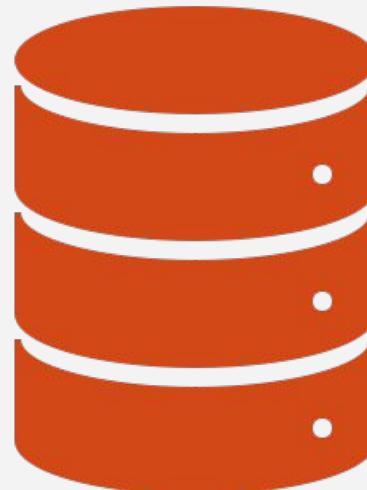
- May be an iterative process
- The data are retrieved
- Checked for completeness
- Descriptive statistics
- Errors corrected, empty fields addressed

Transformation

- Data synchronized (“transformed”) (e.g., M, F, U vs. 1, 2, 9)

Loading

- Data then imported into destination system



5. Check, Clean, and Prepare the Data



Data are now in the system where analysis will be run



Should be a complete set of data



Need to check that everything is ready for analysis



Descriptive statistics

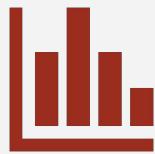


Double-check problem or question being investigated



Double-check against analysis plan

6. Analyze and Interpret the Data



Use the data analysis plan



Perform the actual statistical analyses as described in the plan

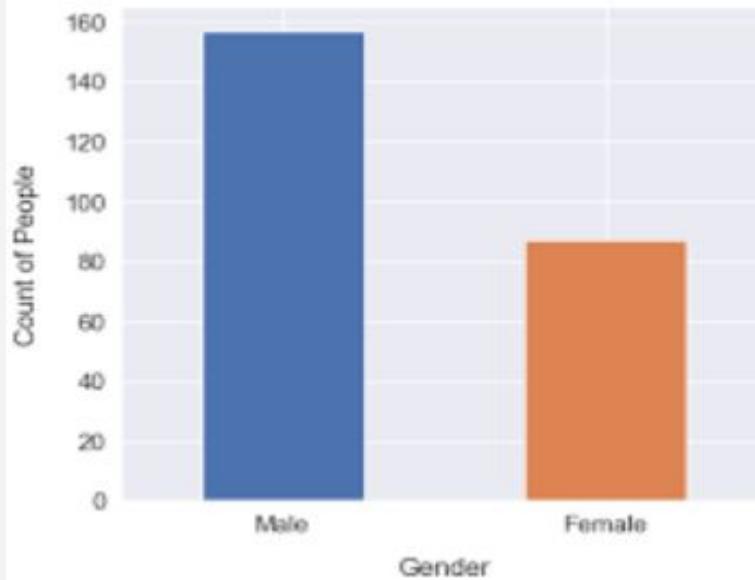


Consult with statistician to confirm interpretations and conclusions

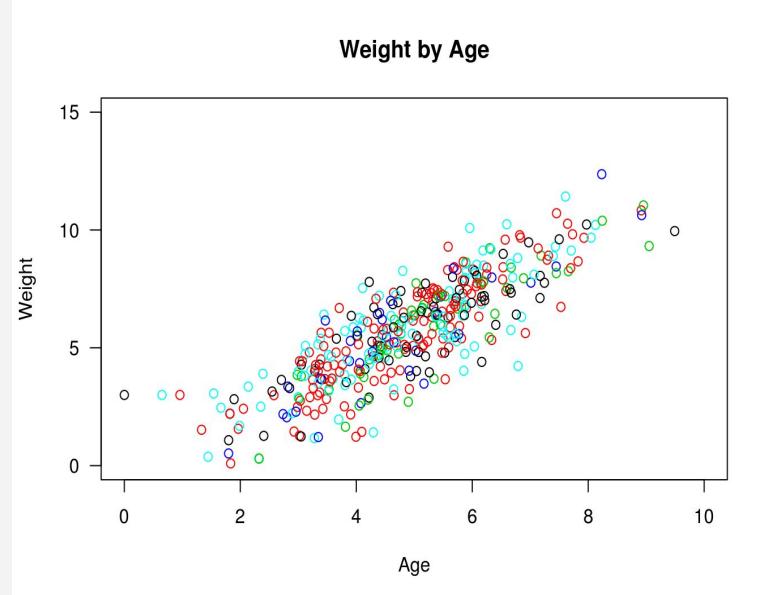
7. Visualize the Data

- **Nominal** (categorical) data: column or bar charts, tables, pie charts, pivot tables
- **Quantitative** data: histograms, scatter plots, line graphs
- Examples of tools
 - Microsoft Excel Chart function
 - PowerBI
 - Tableau
 - R!!

BAR CHART



SCATTER PLOT



8 and 9: Disseminating and Implementing



Disseminating the new knowledge

Write up the findings
Disseminate to the stakeholders



Implementing the new knowledge

Requires participation of stakeholders

In-Class Exercise