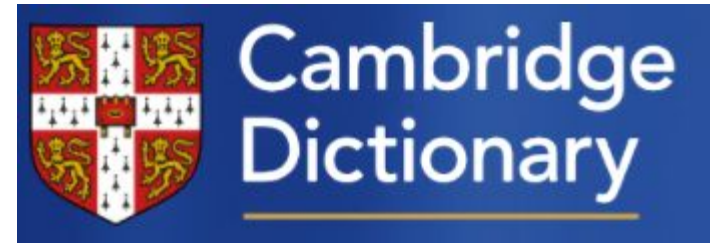


What is data?



The Data Scientist's Toolbox

“Information, especially facts or numbers, collected to be examined and considered and used to help decision making”



“A set of values of qualitative or quantitative variables”

“A set of values of qualitative or quantitative variables”

Set: In statistics, the population you are trying to discover something about

Variable: Measurements or characteristics of an item

Qualitative variable: Measurements or information about qualities

Quantitative variable: Measurements or information about quantities or numerical items



Name	Country of origin	Sex	Weight (kg)	Height (cm)
A. Bee	Canada	M	75	163
C. Dee	UAE	M	80	180
E. Eff	China	F	72	175
G. Haitch	South Africa	F	68	172
I. Jay	Poland	M	77	168
K. Elle	Japan	N/A	76	173
M. Enn	Chile	M	80	190

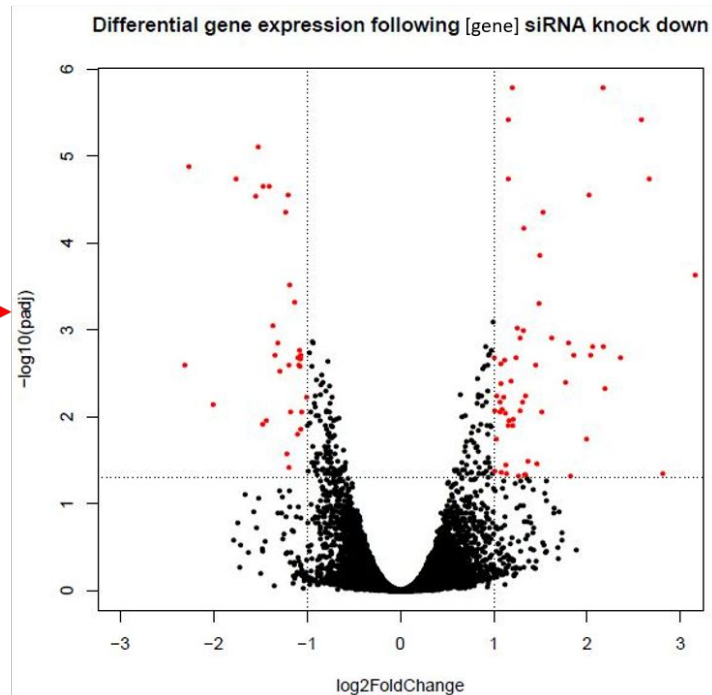
- Sequencing data
- Population census data
- Electronic medical records (EMR), other large databases
- Geographic information system (GIS) data (mapping)
- Image analysis and image extrapolation
- Language and translations
- Website traffic
- Personal/Ad data (eg: Facebook, Netflix predictions, etc)

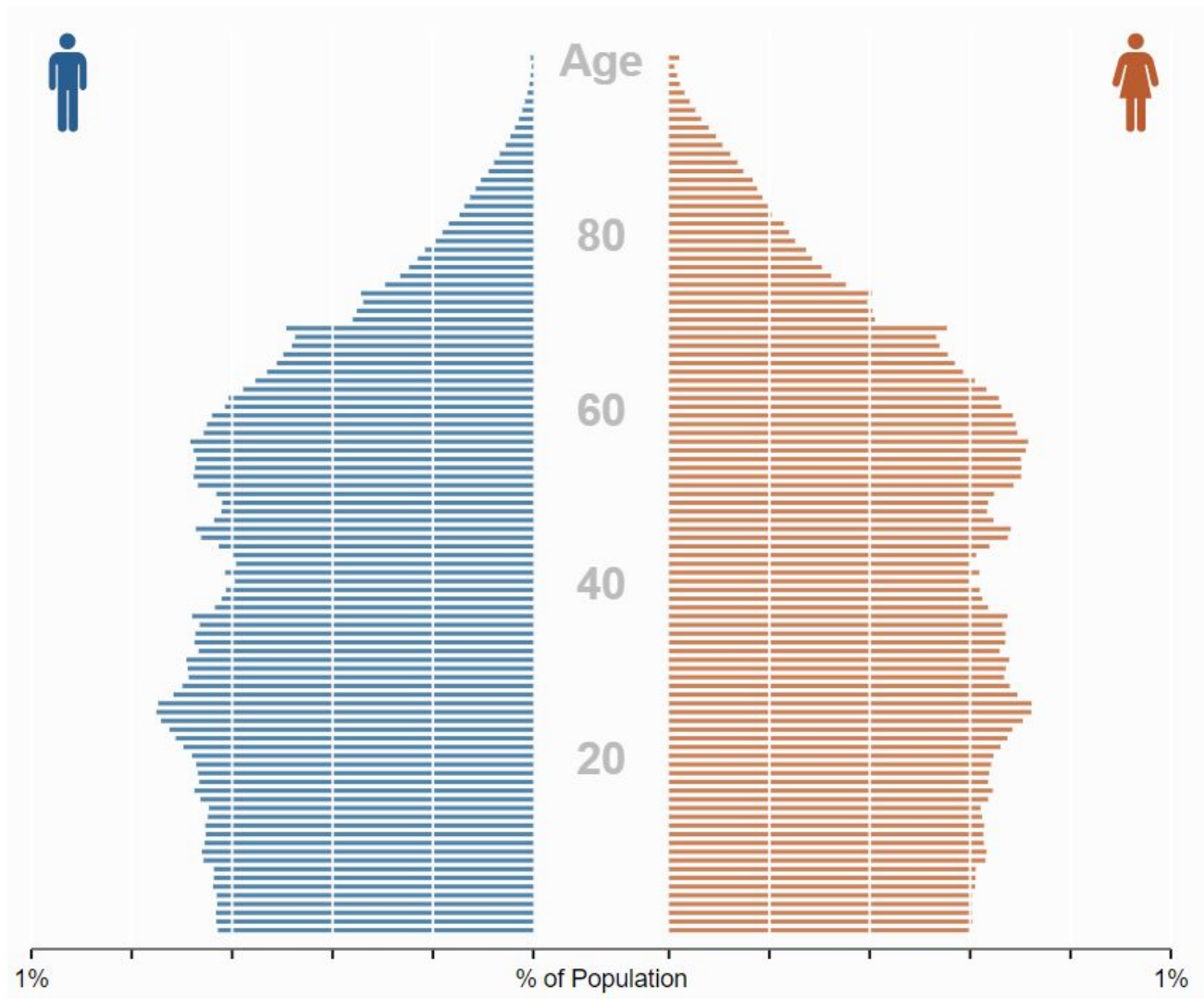



```

CCCCCGGGG9C<6FGAC8E<,8;;C,;<9,EEFFG<F8EEA#:C6@FEGGGGDE#,96C8C#:6##:###:###:
@M01605:159:000000000-B5VR2:1:2113:21912:7245 1:N:0:1
GCCTGGCGGTAGGGGCCAACACCATCTTCCTCACCAAGTTTNTCATGATGATGACANTTCCCCNCTNNTNNTN
+
6@CCCCGGGGEFEGGGGDDGGGG@FGGGDFG9,CFGFGGGFFE#:@CDFGFGCGGA#:0<F@A#:C##:##,#
@M01605:159:000000000-B5VR2:1:2113:17756:7245 1:N:0:1
GACGCTTGCTTAGTGACAGTGACCTACATGGTCACTGAAGANGTGATATAATTTANGATATANATNNTNANNGGN
+
B@-A@EFFC<FFEGGGGG9CEFCA9FFFFFFG9@FFE9@6,#9CC<,<E9,CF<E#:6C<DE#:C##,##:###:C#
@M01605:159:000000000-B5VR2:1:2113:22317:7245 1:N:0:1
GGTGTGCGCTCCTTTTCCCCTGCGGGGCTCCTTTCCCTTCTNACAAATGCTTCACNCTGCTTNCNNCNCNNNGGN
+
CCCCCGAF7+;FFGGGG,<EEFECFG77@,C<9EEF,C,C<#: ,CFGF99CEA<9#6,6,@F#:C##9##,##:,#
@M01605:159:000000000-B5VR2:1:2113:16780:7245 1:N:0:1
AGTCAGCAGGTGTTGGGTCATGTGATCGGGTCTGGCCCATANGTGGCGCTGTTGANCACTCCNGTNNCNCNTNNAAN
+
A@CC@<,CECD,CFG8@FD89<FCG<FGFEFGG,66,BFC#, :C,C+BC+BFF9#6,CFFE#, :##:##6##,9#
@M01605:159:000000000-B5VR2:1:2113:12894:7245 1:N:0:1
ACCATACACTGATCACCTTACTGATAATGCACTGGACACGTNCCTGATAATGCAANTGTCTCNTANNANNANNCAN
+
CCCC8F8,C<FG9,CE8CEA8ECF9FCFFGGGF,CFGFFEC#:6C@9,CFCFF,##,9,6C#,9##,##:###,,#
@M01605:159:000000000-B5VR2:1:2113:21242:7246 1:N:0:1
GACCAGCCCTGGTTGCTGTTGGACATGTTCTGTAGTGTGGGNGGAATACTTGGTCNTTAGCTNCTNNGNNTN
+

```





<ul style="list-style-type: none"> • Help • Logout 	Patient Details <div data-bbox="392 111 566 286">  </div> <div data-bbox="600 142 755 250"> GME0000 Smith, Caroline </div> <div data-bbox="392 304 755 437"> <p>Sex: Female DOB: 1940/01/01 Next of kin: John Smith</p> <p>Phone: 365-565-9090 Address: 19 Provincial Rd. Edmonton AB T6M 1R7</p> </div>			GP Details <p>Name: Jones, Evans Phone: 333-465-5545 Address: 11 Terrence Ave., Edmonton, AB T4Y 8U9</p>																																																																																																																																																																																									
Patient Record <ul style="list-style-type: none"> • Summary • Lab Results • Diagnostic Images • Details • Notes or Comments 	Alerts <p>Allergies - Sulfa Drugs</p> <ul style="list-style-type: none"> • Pap smear due • Td due • A1C above target 			Other Healthcare Providers <table border="1"> <thead> <tr> <th>Name</th> <th>Disp.</th> <th>Last Encounter</th> <th>Next encounter</th> <th>Right of Access</th> </tr> </thead> <tbody> <tr> <td>Diaz, Ellen</td> <td>Cardiology</td> <td>01/2006</td> <td>07/2006</td> <td>Y</td> </tr> <tr> <td>Fournier, Janice</td> <td>RN</td> <td>08/2005</td> <td></td> <td>N</td> </tr> <tr> <td>Cohen, Richard</td> <td>Dermatology</td> <td>07/2005</td> <td></td> <td>N</td> </tr> </tbody> </table>						Name	Disp.	Last Encounter	Next encounter	Right of Access	Diaz, Ellen	Cardiology	01/2006	07/2006	Y	Fournier, Janice	RN	08/2005		N	Cohen, Richard	Dermatology	07/2005		N																																																																																																																																																																
Name	Disp.	Last Encounter	Next encounter	Right of Access																																																																																																																																																																																									
Diaz, Ellen	Cardiology	01/2006	07/2006	Y																																																																																																																																																																																									
Fournier, Janice	RN	08/2005		N																																																																																																																																																																																									
Cohen, Richard	Dermatology	07/2005		N																																																																																																																																																																																									
<table border="1"> <thead> <tr> <th>Diagnosis</th> <th>Date</th> <th>Status</th> </tr> </thead> <tbody> <tr> <td>Hypertension</td> <td>11/1989</td> <td>Ongoing</td> </tr> <tr> <td>Diabetes</td> <td>05/1996</td> <td>Ongoing</td> </tr> <tr> <td>Coronary Artery Disease</td> <td>02/2002</td> <td>Ongoing</td> </tr> <tr> <td>Fasting lipids</td> <td>12/2005</td> <td></td> </tr> <tr> <td>Exercise stress test</td> <td>1/2005</td> <td></td> </tr> <tr> <td>Coronary angiogram / Cellulitis</td> <td>02/2005</td> <td>Resolved</td> </tr> <tr> <td>Cholecystectomy</td> <td>05/1981</td> <td>Resolved</td> </tr> <tr> <td>Cesarian section</td> <td>01/1967</td> <td>Resolved</td> </tr> </tbody> </table>			Diagnosis	Date	Status	Hypertension	11/1989	Ongoing	Diabetes	05/1996	Ongoing	Coronary Artery Disease	02/2002	Ongoing	Fasting lipids	12/2005		Exercise stress test	1/2005		Coronary angiogram / Cellulitis	02/2005	Resolved	Cholecystectomy	05/1981	Resolved	Cesarian section	01/1967	Resolved	Medications <table border="1"> <thead> <tr> <th>Date</th> <th>Medications</th> <th>Prescriptions</th> <th>Last Filled</th> </tr> </thead> <tbody> <tr> <td>11/1989</td> <td>Hydrochlorothiazide 25 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>03/1999</td> <td>Glyburide 5 mg</td> <td>One tab twice daily</td> <td>12/2005</td> </tr> <tr> <td>01/2001</td> <td>Metformin 500 mg</td> <td>Two tabs twice daily</td> <td>12/2005</td> </tr> <tr> <td>03/2001</td> <td>Atorvastatin 20 mg</td> <td>One tab at supper</td> <td>12/2005</td> </tr> <tr> <td>02/2002</td> <td>Atenolol 50 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>02/2002</td> <td>ECASA 325 mg</td> <td>One tab at breakfast</td> <td>12/2005</td> </tr> <tr> <td>02/2006</td> <td>Ramipril 10mg</td> <td>One tab at supper</td> <td>02/2006</td> </tr> <tr> <td>06/2005</td> <td>Cloxacillin 500 mg</td> <td>Discontinued</td> <td></td> </tr> <tr> <td>05/2004</td> <td>Beclomethasone Cream</td> <td>Discontinued</td> <td></td> </tr> </tbody> </table> Encounter History <table border="1"> <thead> <tr> <th>Date</th> <th>Facility</th> <th>Speciality</th> <th>Clinician</th> <th>Reason</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>02/2006</td> <td>GP</td> <td></td> <td></td> <td>Hypertension</td> <td>-</td> </tr> <tr> <td>01/2006</td> <td>Cardio Assoc</td> <td>Cardiology</td> <td>Diaz, E.</td> <td>CAD</td> <td>Outpatient</td> </tr> <tr> <td>12/2005</td> <td>GP</td> <td></td> <td></td> <td>Diabetes</td> <td>-</td> </tr> <tr> <td>10/2005</td> <td>General Hosp</td> <td>Dietician</td> <td>Johnson, H.</td> <td>Diabetes teaching</td> <td>Outpatient</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Diabetes</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>Home Visit</td> <td>RN</td> <td>Fournier, J.</td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>08/2005</td> <td>GP</td> <td></td> <td></td> <td>Cellulitis</td> <td>-</td> </tr> <tr> <td>07/2005</td> <td>Polyclinic</td> <td>Dermatology</td> <td>Cohen, R.</td> <td>Stasis dermatitis</td> <td>Outpatient</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="3">Immunizations</th> <th colspan="3">Diabetic Indices</th> </tr> <tr> <th>Type</th> <th>Most Recent</th> <th>Number Received</th> <th>Type</th> <th>Value</th> <th>Most Recent</th> </tr> </thead> <tbody> <tr> <td>Influenza</td> <td>11/2005</td> <td>7</td> <td>A1C</td> <td>0.071</td> <td>12/2005</td> </tr> <tr> <td>Pneumovax</td> <td>03/2005</td> <td>1</td> <td>LDL</td> <td>2.41</td> <td>12/2005</td> </tr> <tr> <td>Twinrix</td> <td>08/2002</td> <td>3</td> <td>BP</td> <td>135/75</td> <td>02/2006</td> </tr> <tr> <td>Td</td> <td>04/1996</td> <td>1</td> <td>Urine</td> <td></td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td>Microalb</td> <td>0.02</td> <td>08/2005</td> </tr> <tr> <td></td> <td></td> <td></td> <td>Eye Exam</td> <td></td> <td>05/2005</td> </tr> <tr> <td></td> <td></td> <td></td> <td>Home Gluc (average)</td> <td>7.4</td> <td>01/2006</td> </tr> </tbody> </table>						Date	Medications	Prescriptions	Last Filled	11/1989	Hydrochlorothiazide 25 mg	One tab at breakfast	12/2005	03/1999	Glyburide 5 mg	One tab twice daily	12/2005	01/2001	Metformin 500 mg	Two tabs twice daily	12/2005	03/2001	Atorvastatin 20 mg	One tab at supper	12/2005	02/2002	Atenolol 50 mg	One tab at breakfast	12/2005	02/2002	ECASA 325 mg	One tab at breakfast	12/2005	02/2006	Ramipril 10mg	One tab at supper	02/2006	06/2005	Cloxacillin 500 mg	Discontinued		05/2004	Beclomethasone Cream	Discontinued		Date	Facility	Speciality	Clinician	Reason	Type	02/2006	GP			Hypertension	-	01/2006	Cardio Assoc	Cardiology	Diaz, E.	CAD	Outpatient	12/2005	GP			Diabetes	-	10/2005	General Hosp	Dietician	Johnson, H.	Diabetes teaching	Outpatient	08/2005	GP			Diabetes	-	08/2005	GP			Cellulitis	-	08/2005	Home Visit	RN	Fournier, J.	Cellulitis	-	08/2005	GP			Cellulitis	-	07/2005	Polyclinic	Dermatology	Cohen, R.	Stasis dermatitis	Outpatient	Immunizations			Diabetic Indices			Type	Most Recent	Number Received	Type	Value	Most Recent	Influenza	11/2005	7	A1C	0.071	12/2005	Pneumovax	03/2005	1	LDL	2.41	12/2005	Twinrix	08/2002	3	BP	135/75	02/2006	Td	04/1996	1	Urine						Microalb	0.02	08/2005				Eye Exam		05/2005				Home Gluc (average)	7.4	01/2006
Diagnosis	Date	Status																																																																																																																																																																																											
Hypertension	11/1989	Ongoing																																																																																																																																																																																											
Diabetes	05/1996	Ongoing																																																																																																																																																																																											
Coronary Artery Disease	02/2002	Ongoing																																																																																																																																																																																											
Fasting lipids	12/2005																																																																																																																																																																																												
Exercise stress test	1/2005																																																																																																																																																																																												
Coronary angiogram / Cellulitis	02/2005	Resolved																																																																																																																																																																																											
Cholecystectomy	05/1981	Resolved																																																																																																																																																																																											
Cesarian section	01/1967	Resolved																																																																																																																																																																																											
Date	Medications	Prescriptions	Last Filled																																																																																																																																																																																										
11/1989	Hydrochlorothiazide 25 mg	One tab at breakfast	12/2005																																																																																																																																																																																										
03/1999	Glyburide 5 mg	One tab twice daily	12/2005																																																																																																																																																																																										
01/2001	Metformin 500 mg	Two tabs twice daily	12/2005																																																																																																																																																																																										
03/2001	Atorvastatin 20 mg	One tab at supper	12/2005																																																																																																																																																																																										
02/2002	Atenolol 50 mg	One tab at breakfast	12/2005																																																																																																																																																																																										
02/2002	ECASA 325 mg	One tab at breakfast	12/2005																																																																																																																																																																																										
02/2006	Ramipril 10mg	One tab at supper	02/2006																																																																																																																																																																																										
06/2005	Cloxacillin 500 mg	Discontinued																																																																																																																																																																																											
05/2004	Beclomethasone Cream	Discontinued																																																																																																																																																																																											
Date	Facility	Speciality	Clinician	Reason	Type																																																																																																																																																																																								
02/2006	GP			Hypertension	-																																																																																																																																																																																								
01/2006	Cardio Assoc	Cardiology	Diaz, E.	CAD	Outpatient																																																																																																																																																																																								
12/2005	GP			Diabetes	-																																																																																																																																																																																								
10/2005	General Hosp	Dietician	Johnson, H.	Diabetes teaching	Outpatient																																																																																																																																																																																								
08/2005	GP			Diabetes	-																																																																																																																																																																																								
08/2005	GP			Cellulitis	-																																																																																																																																																																																								
08/2005	Home Visit	RN	Fournier, J.	Cellulitis	-																																																																																																																																																																																								
08/2005	GP			Cellulitis	-																																																																																																																																																																																								
07/2005	Polyclinic	Dermatology	Cohen, R.	Stasis dermatitis	Outpatient																																																																																																																																																																																								
Immunizations			Diabetic Indices																																																																																																																																																																																										
Type	Most Recent	Number Received	Type	Value	Most Recent																																																																																																																																																																																								
Influenza	11/2005	7	A1C	0.071	12/2005																																																																																																																																																																																								
Pneumovax	03/2005	1	LDL	2.41	12/2005																																																																																																																																																																																								
Twinrix	08/2002	3	BP	135/75	02/2006																																																																																																																																																																																								
Td	04/1996	1	Urine																																																																																																																																																																																										
			Microalb	0.02	08/2005																																																																																																																																																																																								
			Eye Exam		05/2005																																																																																																																																																																																								
			Home Gluc (average)	7.4	01/2006																																																																																																																																																																																								



QUESTION



DATA



Summarizing: What is data?



The Data Scientist's Toolbox