Data and Variables

Gaston Sanchez

Statistics is...

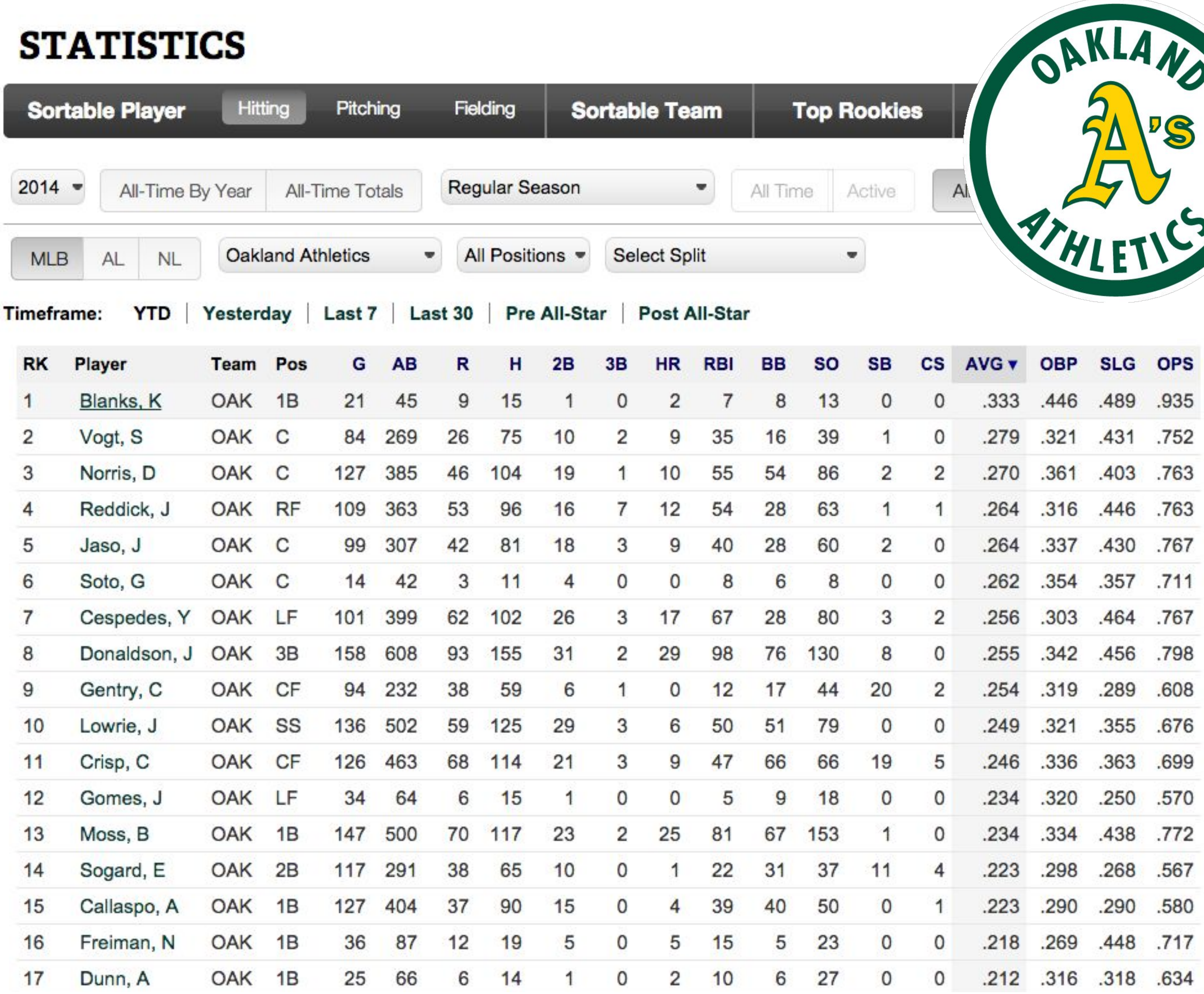
Ordinary use

In ordinary conversations, the word "statistics" is used as a term to indicate a set or collection of numeric records

Common example

Baseball Statistics

(or any other sports statistics)



STATISTICS

2014 | Regular Season | Oakland Athletics

Timeframe: YTD | Yesterday | Last 7 | Last 30 | Pre All-Star | Post All-Star

RK	Player	Team	Pos	G	AB	R	H	2B	3B	HR	RBI	BB	SO	SB	CS	AVG ▾	OBP	SLG	OPS
1	Blanks, K	OAK	1B	21	45	9	15	1	0	2	7	8	13	0	0	.333	.446	.489	.935
2	Vogt, S	OAK	C	84	269	26	75	10	2	9	35	16	39	1	0	.279	.321	.431	.752
3	Norris, D	OAK	C	127	385	46	104	19	1	10	55	54	86	2	2	.270	.361	.403	.763
4	Reddick, J	OAK	RF	109	363	53	96	16	7	12	54	28	63	1	1	.264	.316	.446	.763
5	Jaso, J	OAK	C	99	307	42	81	18	3	9	40	28	60	2	0	.264	.337	.430	.767
6	Soto, G	OAK	C	14	42	3	11	4	0	0	8	6	8	0	0	.262	.354	.357	.711
7	Cespedes, Y	OAK	LF	101	399	62	102	26	3	17	67	28	80	3	2	.256	.303	.464	.767
8	Donaldson, J	OAK	3B	158	608	93	155	31	2	29	98	76	130	8	0	.255	.342	.456	.798
9	Gentry, C	OAK	CF	94	232	38	59	6	1	0	12	17	44	20	2	.254	.319	.289	.608
10	Lowrie, J	OAK	SS	136	502	59	125	29	3	6	50	51	79	0	0	.249	.321	.355	.676
11	Crisp, C	OAK	CF	126	463	68	114	21	3	9	47	66	66	19	5	.246	.336	.363	.699
12	Gomes, J	OAK	LF	34	64	6	15	1	0	0	5	9	18	0	0	.234	.320	.250	.570
13	Moss, B	OAK	1B	147	500	70	117	23	2	25	81	67	153	1	0	.234	.334	.438	.772
14	Sogard, E	OAK	2B	117	291	38	65	10	0	1	22	31	37	11	4	.223	.298	.268	.567
15	Callaspo, A	OAK	1B	127	404	37	90	15	0	4	39	40	50	0	1	.223	.290	.290	.580
16	Freiman, N	OAK	1B	36	87	12	19	5	0	5	15	5	23	0	0	.218	.269	.448	.717
17	Dunn, A	OAK	1B	25	66	6	14	1	0	2	10	6	27	0	0	.212	.316	.318	.634

Origins of the term Statistics

From German **Statistik**

Coined by Gottfried Achenwall (1749)

Science of State: analysis of data about the State

"Political Arithmetic" (in English)

Data used by the government; Census; National Statistics Institutes

Statistics as a discipline

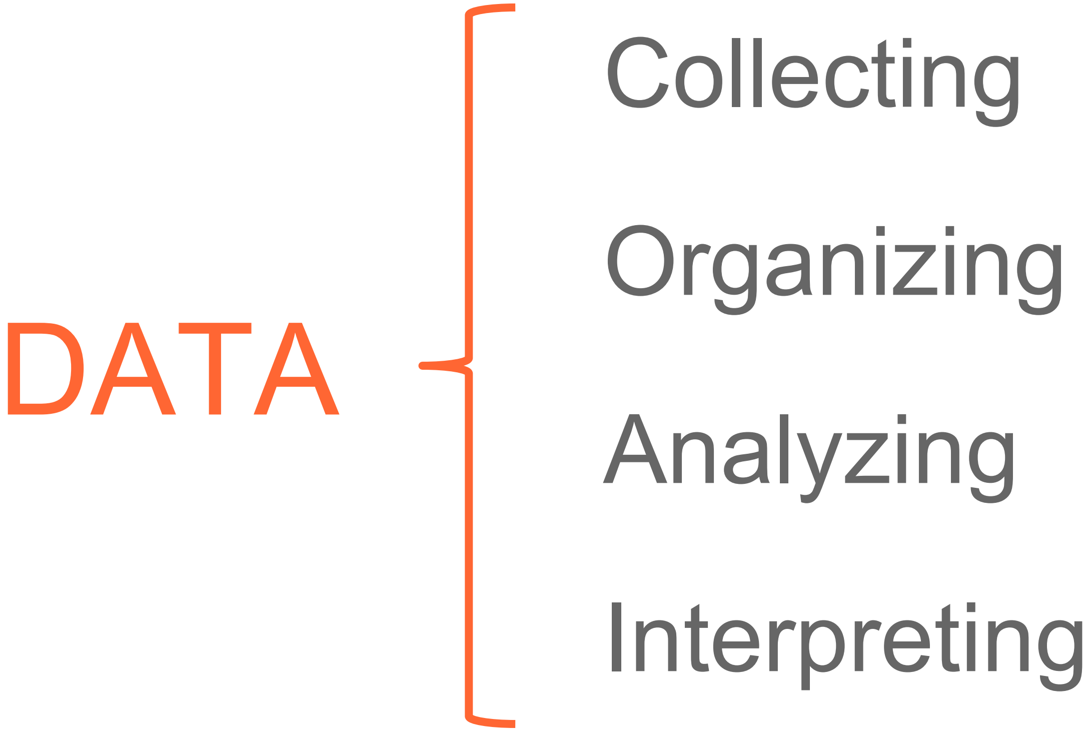
We are concerned about "Statistics" in a broader formal sense; as an analytical discipline

Statistics
The Science of Data

“Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.”

WIKIPEDIA
The Free Encyclopedia

Statistics



DATA

Collecting

Organizing

Analyzing

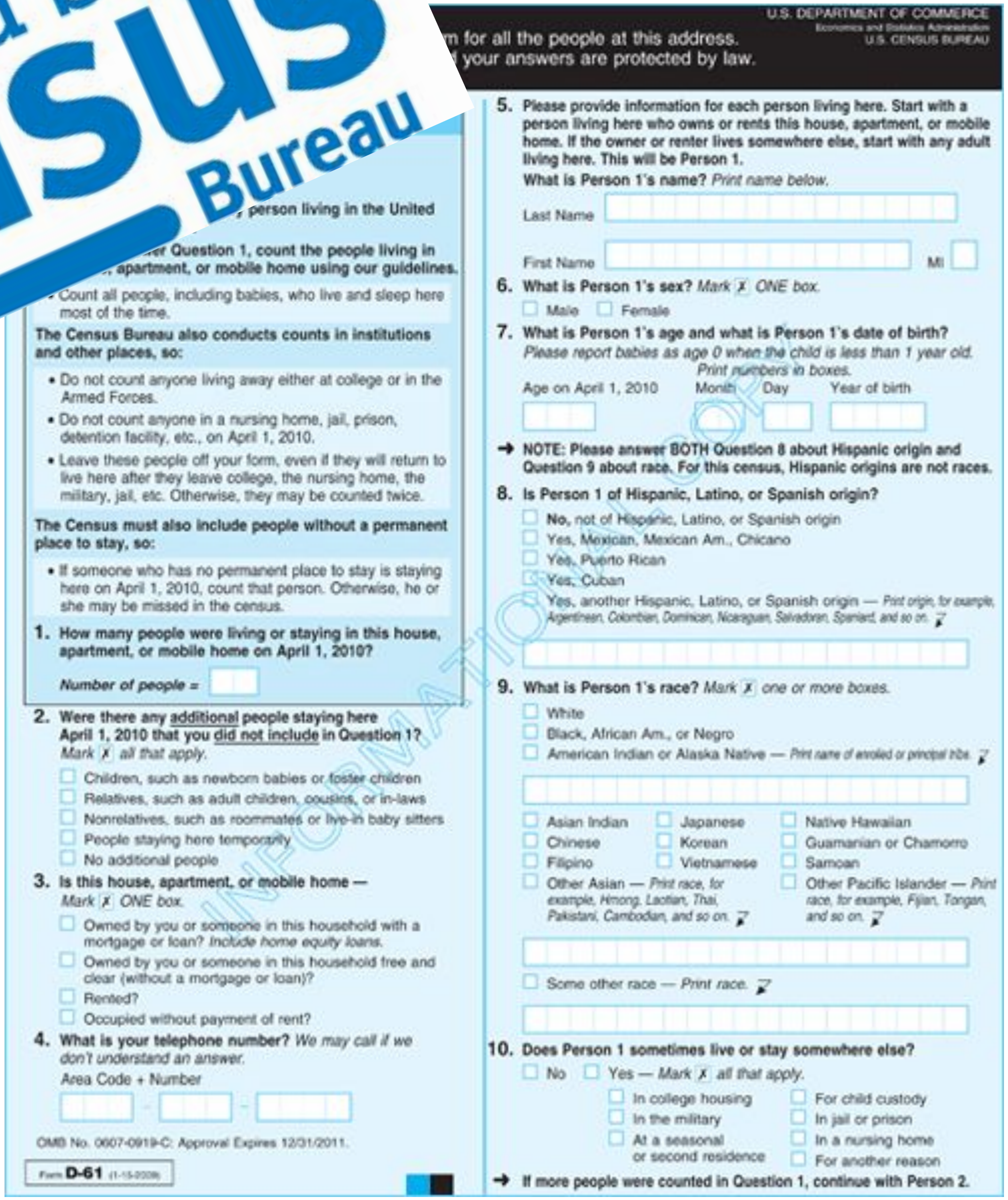
Interpreting

DATA?

Sources of Data

SURVEY
VERY SATISFIED
ATISFIED
3
4
5
6
7
8
9

United States™
Census
Bureau



m for all the people at this address.
l your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

y person living in the United

er Question 1, count the people living in
, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people = □□

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark ☒ all that apply.*

□ Children, such as newborn babies or foster children
□ Relatives, such as adult children, cousins, or in-laws
□ Nonrelatives, such as roommates or live-in baby sitters
□ People staying here temporarily
□ No additional people

3. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

□ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
□ Owned by you or someone in this household free and clear (without a mortgage or loan)?
□ Rented?
□ Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

□□□ - □□□ - □□□□

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (1-15-2009)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name

First Name MI

6. What is Person 1's sex? *Mark ☒ ONE box.*

□ Male □ Female

7. What is Person 1's age and what is Person 1's date of birth?
Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.

Age on April 1, 2010 Month Day Year of birth

→ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

□ **No,** not of Hispanic, Latino, or Spanish origin
□ Yes, Mexican, Mexican Am., Chicano
□ Yes, Puerto Rican
□ Yes, Cuban
□ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark ☒ one or more boxes.*

□ White
□ Black, African Am., or Negro
□ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

□ Asian Indian
□ Chinese
□ Filipino
□ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
□ Japanese
□ Korean
□ Vietnamese
□ Native Hawaiian
□ Guamanian or Chamorro
□ Samoan
□ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

□ Some other race — *Print race.*

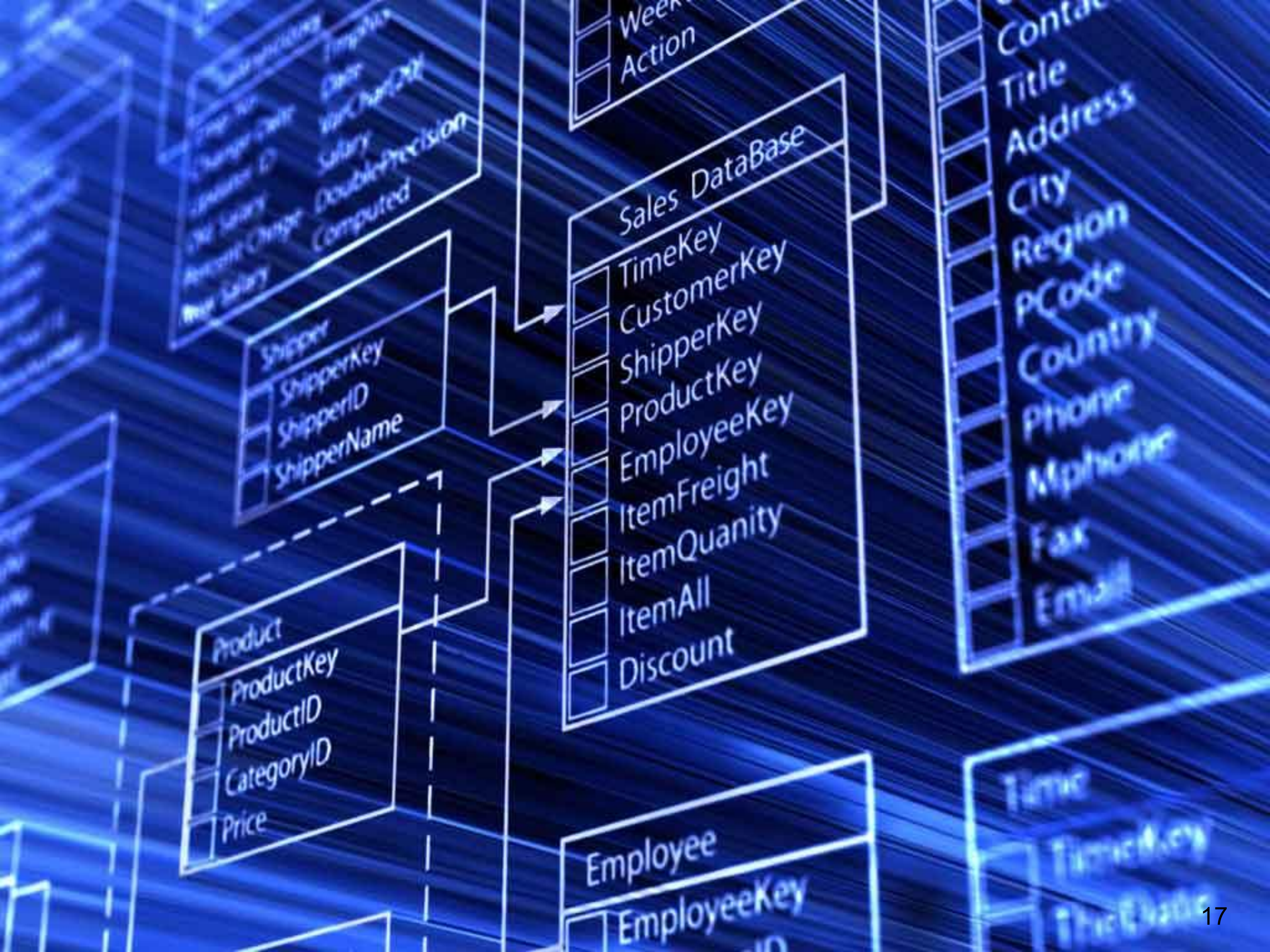
10. Does Person 1 sometimes live or stay somewhere else?

□ No □ Yes — *Mark ☒ all that apply.*

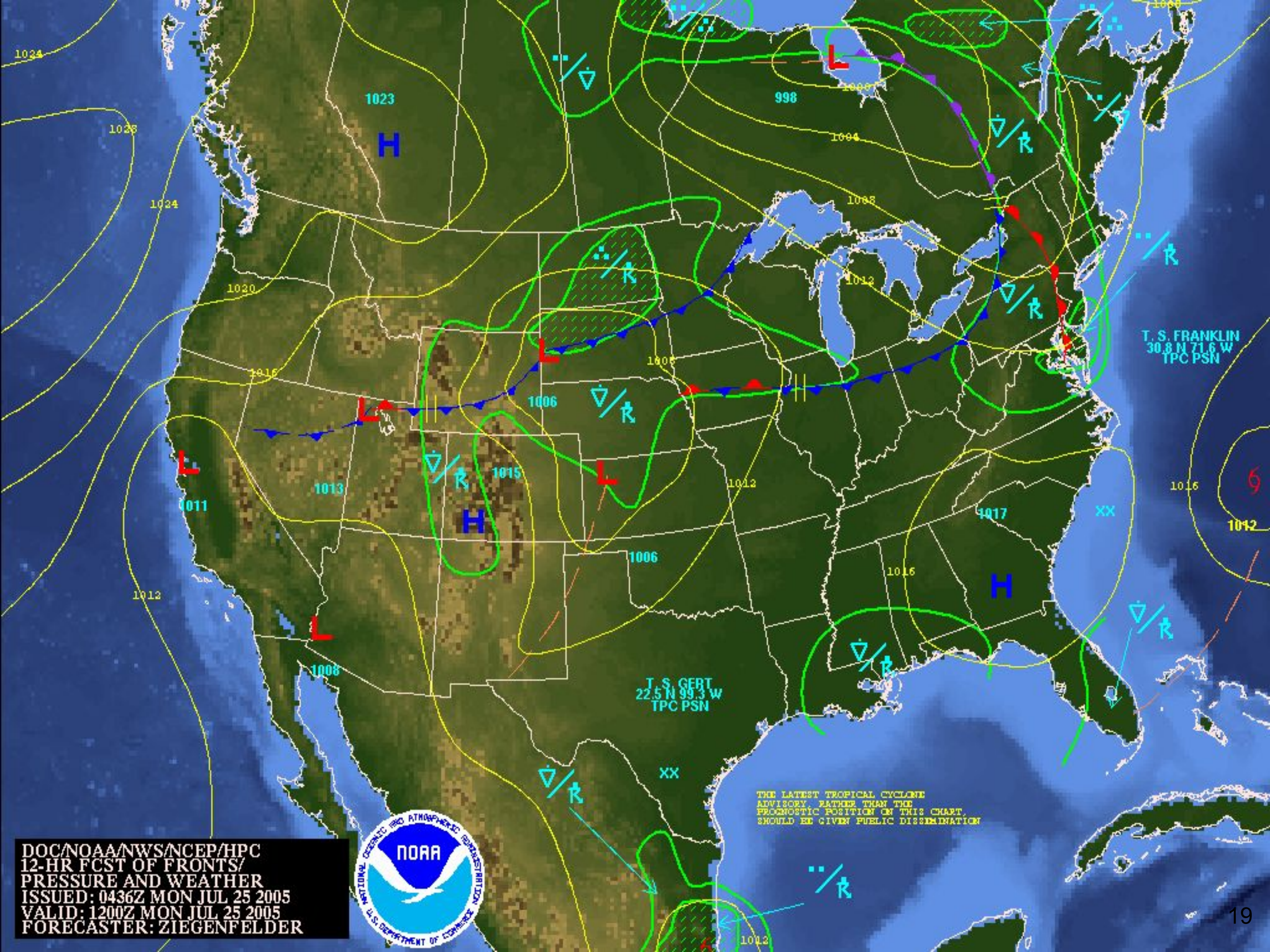
□ In college housing
□ In the military
□ At a seasonal or second residence
□ For child custody
□ In jail or prison
□ In a nursing home
□ For another reason

→ **If more people were counted in Question 1, continue with Person 2.**

0910
Flag Ref
Abnormal
ted: 01
K/uL
M/uL
gm/dL
fL
pg
mg/dL
PAGE
** CONTINUED ON


Action
Sales DataBase
TimeKey
CustomerKey
ShipperKey
ProductKey
EmployeeKey
ItemFreight
ItemQuanity
ItemAll
Discount
Title
Address
City
Region
PCode
Country
Phone
Mphone
Fax
Email
Shipper
ShipperKey
ShipperID
ShipperName
Product
ProductKey
ProductID
CategoryID
Price
Employee
EmployeeKey
Time
TimeKey
TheDate
Salary
DoublePrecision
Computed

Settings
11:23 AM
Mission Minis
Charge
$1.00
Enter an optional note
Clear
7
8
9
4
5
6
1
2
3
00
0


DOC/NOAA/NWS/NCEP/HPC
12-HR FCST OF FRONTS/
PRESSURE AND WEATHER
ISSUED: 0436Z MON JUL 25 2005
VALID: 1200Z MON JUL 25 2005
FORECASTER: ZIEGENFELDER
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
THE LATEST TROPICAL CYCLONE ADVISORY, RATHER THAN THE PROGNOSTIC POSITION ON THIS CHART, SHOULD BE GIVEN PUBLIC DISSEMINATION
T. S. FRANKLIN
30.8 N 71.6 W
TPC PSN
T. S. GERT
22.5 N 99.3 W
TPC PSN
H
L
1023
998
1006
1015
1013
1011
1008
1017
1012
1016
1020
1024
1028
1000
1004
XX

LinkedIn
skype
LiVEJOU
foursquare
g+
vime
Insta
YouTube
facebook
g+
Blogger
my
nstagram
twitter
flickr
tumblr.
yelp
Lik

Data for Statistical Analysis

The raw material of Statistics is Data

Data in Statistics

In Statistics, “Data” is often conceptualized as having a set of **objects** on which we observe or measure one or more **characteristics**

Some Terminology

Objects & Characteristics

individuals **variables**

Why “Variable”?

A characteristic that **varies** from one individual to another

Some Terminology

individuals

observations

subjects

objects

cases

variables

characteristics

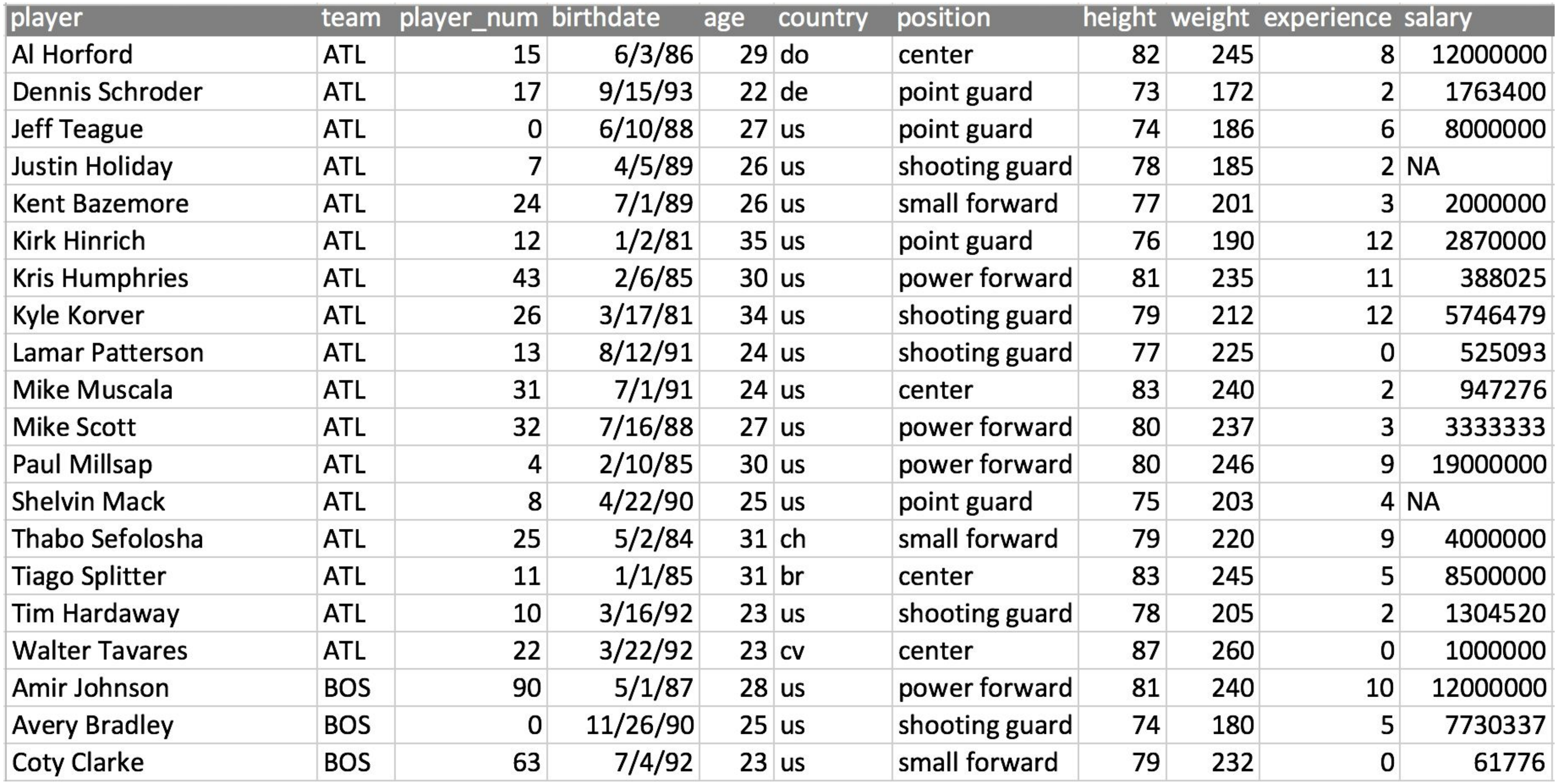
attributes

features

traits

Example

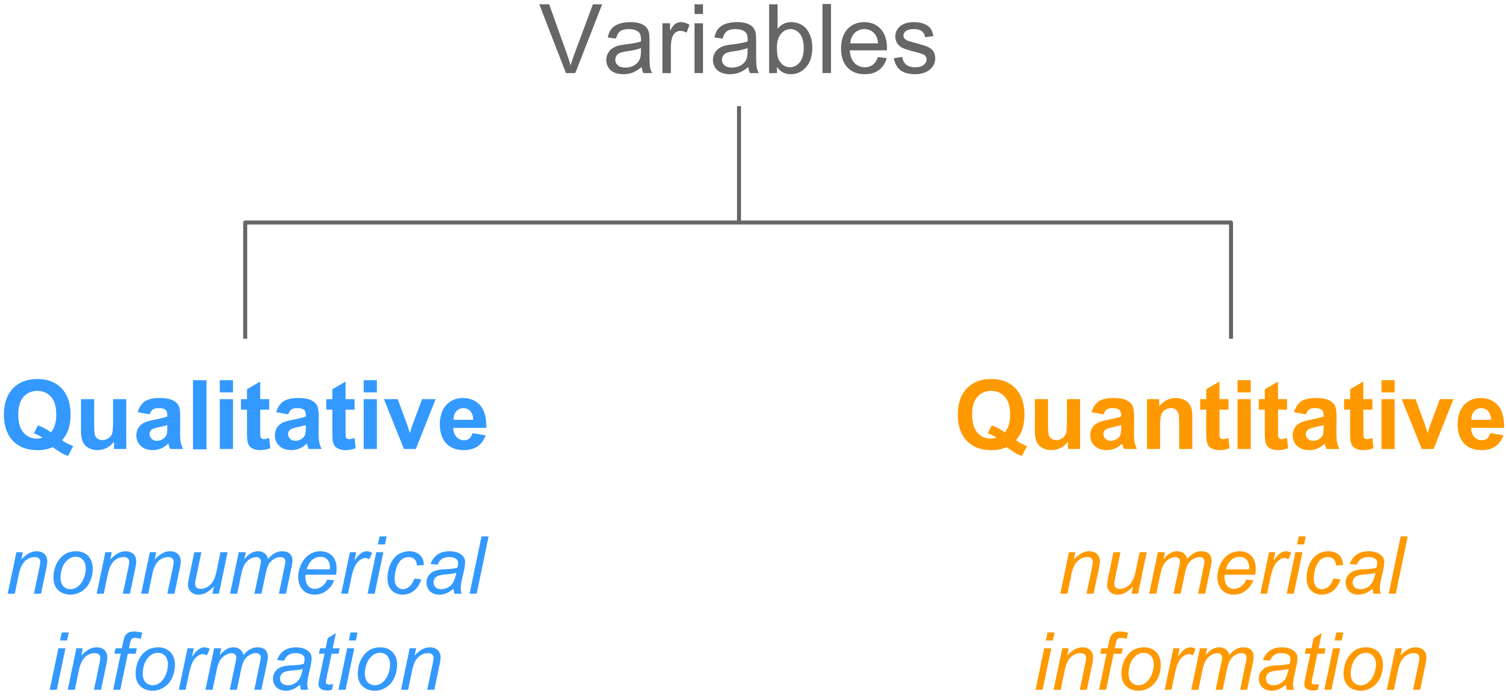
NEW YORK KNICKS
CLEVELAND CAVALIERS
CHICAGO BULLS
DETROIT PISTONS
ATLANTA HAWKS
BOSTON CELTICS
TORONTO RAPTORS
DENVER NUGGETS
ORLANDO MAGIC
LOS ANGELES CLIPPERS
HOUSTON ROCKETS
GOLDEN STATE WARRIORS
MINNESOTA TIMBERWOLVES
PORTLAND TRAIL BLAZERS
PHOENIX SUNS
WIZARDS
THUNDER
OKC
MIAMI HEAT
NBA
NEW ORLEANS HORNETS
NOLA
CHARLOTTE BOBCATS
LOS ANGELES LAKERS
Pacers
JAZZ
NETS
B
BROOKLYN
MILWAUKEE BUCKS
SACRAMENTO KINGS
MEMPHIS GRIZZLIES
SAN ANTONIO SPURS
DALLAS MAVERICKS
76ers
PHILADELPHIA



player	team	player_num	birthdate	age	country	position	height	weight	experience	salary
Al Horford	ATL	15	6/3/86	29	do	center	82	245	8	12000000
Dennis Schroder	ATL	17	9/15/93	22	de	point guard	73	172	2	1763400
Jeff Teague	ATL	0	6/10/88	27	us	point guard	74	186	6	8000000
Justin Holiday	ATL	7	4/5/89	26	us	shooting guard	78	185	2	NA
Kent Bazemore	ATL	24	7/1/89	26	us	small forward	77	201	3	2000000
Kirk Hinrich	ATL	12	1/2/81	35	us	point guard	76	190	12	2870000
Kris Humphries	ATL	43	2/6/85	30	us	power forward	81	235	11	388025
Kyle Korver	ATL	26	3/17/81	34	us	shooting guard	79	212	12	5746479
Lamar Patterson	ATL	13	8/12/91	24	us	shooting guard	77	225	0	525093
Mike Muscala	ATL	31	7/1/91	24	us	center	83	240	2	947276
Mike Scott	ATL	32	7/16/88	27	us	power forward	80	237	3	3333333
Paul Millsap	ATL	4	2/10/85	30	us	power forward	80	246	9	19000000
Shelvin Mack	ATL	8	4/22/90	25	us	point guard	75	203	4	NA
Thabo Sefolosha	ATL	25	5/2/84	31	ch	small forward	79	220	9	4000000
Tiago Splitter	ATL	11	1/1/85	31	br	center	83	245	5	8500000
Tim Hardaway	ATL	10	3/16/92	23	us	shooting guard	78	205	2	1304520
Walter Tavares	ATL	22	3/22/92	23	cv	center	87	260	0	1000000
Amir Johnson	BOS	90	5/1/87	28	us	power forward	81	240	10	12000000
Avery Bradley	BOS	0	11/26/90	25	us	shooting guard	74	180	5	7730337
Coty Clarke	BOS	63	7/4/92	23	us	small forward	79	232	0	61776

Variables

Variables play the starring role in statistical studies



Variables

Qualitative

nonnumerical information

Quantitative

numerical information

Some qualitative variables

Team

ATL, BOS, GSW

Position

Center, Point Guard, Shooting Guard

Country

USA, Brazil, Canada, Australia

Some quantitative variables

Age (yrs)

29, 22, 27, 26

Height (in)

82, 73, 74, 78

Salary (millions of dls)

12, 1.7, 8, 0.38,

What about ...

What type of variables are these?

Player Name

Player Number

Birthdate

Team Ranking

When numbers are used to codify qualities ...

Assigning numbers to qualities

Gender of newborn

male = 0, female = 1

Icecream Flavors

chocolate = 10

vanilla = 20

lemon = 30

Assigning numbers to qualities

Frequency of usage

never = 0

rarely = 1

sometimes = 2

often = 3

always = 4

Think about it ...

Assigning numbers to qualities

Icecream Flavors

chocolate = 10

vanilla = 20

lemon = 30

30 - 20 = 10?

(lemon - vanilla) = chocolate?

Assigning numbers to qualities

Frequency of usage

never = 0

rarely = 1

sometimes = 2

often = 3

always = 4

Is always (4) twice as sometimes (2)?

Assigning numbers to qualities

0	2	1	4	4	3
never	some	rarely	always	always	often

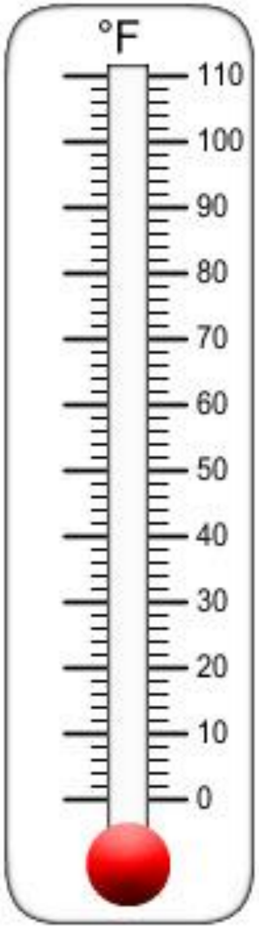
Avg = 2.3

What does it mean?

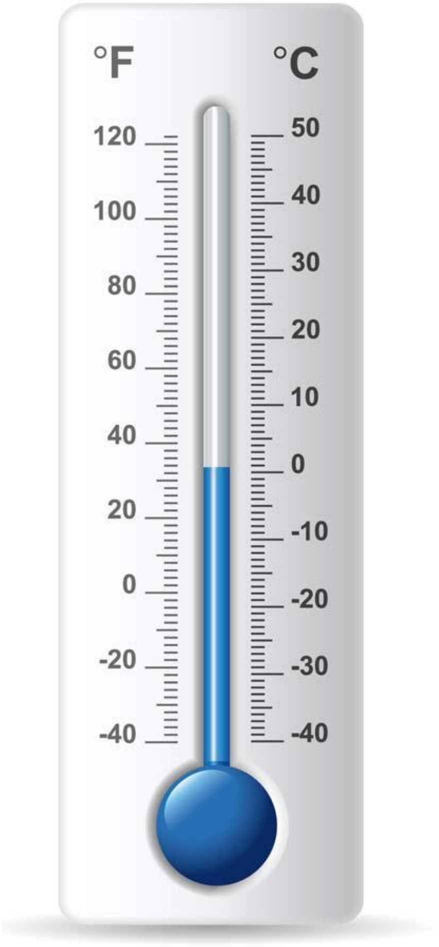
Discussion

How would you change a **quantitative** variable into a **qualitative** one?

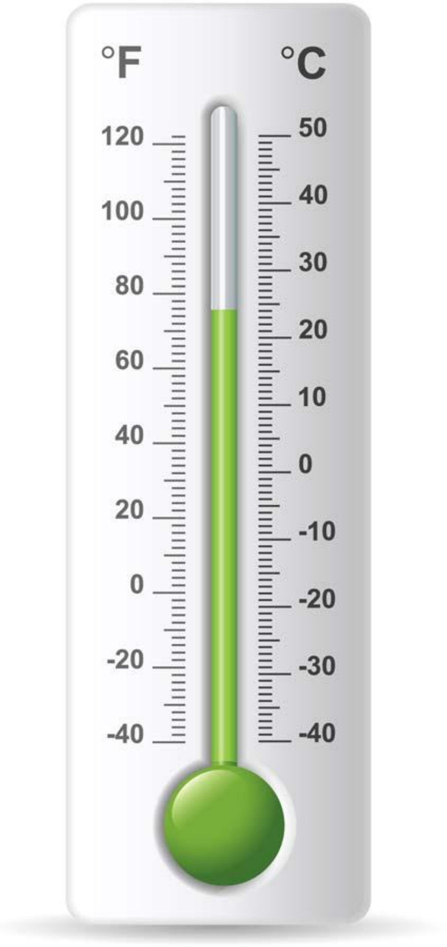
Converting Temperature into a qualitative variable



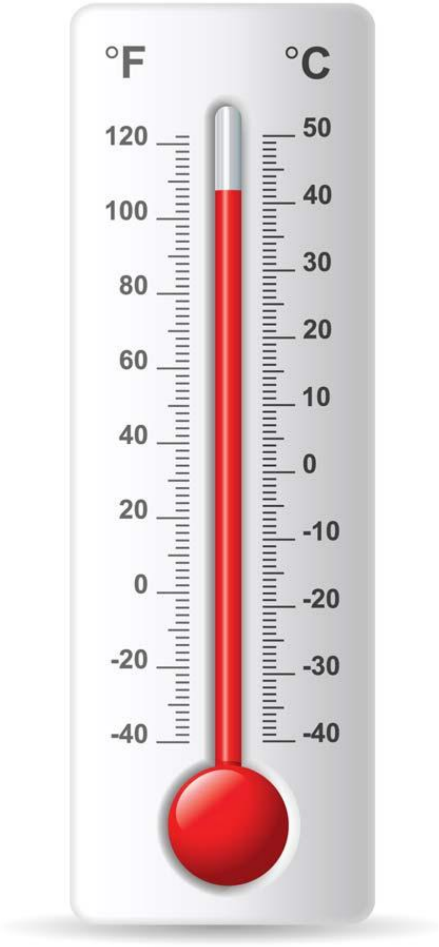
Cold
< 5 ^{o}C



Mild
5 - 25 ^{o}C



Hot
> 25 ^{o}C



How would you change a **qualitative** variable into a **quantitative** one?

Switching to quantitative variables

Frequency of usage

never
rarely
sometimes
often
always

Quantifying

times / day
# days
# weeks
months
years

Switching to quantitative variables

Icecream flavors

chocolate

vanilla

lemon

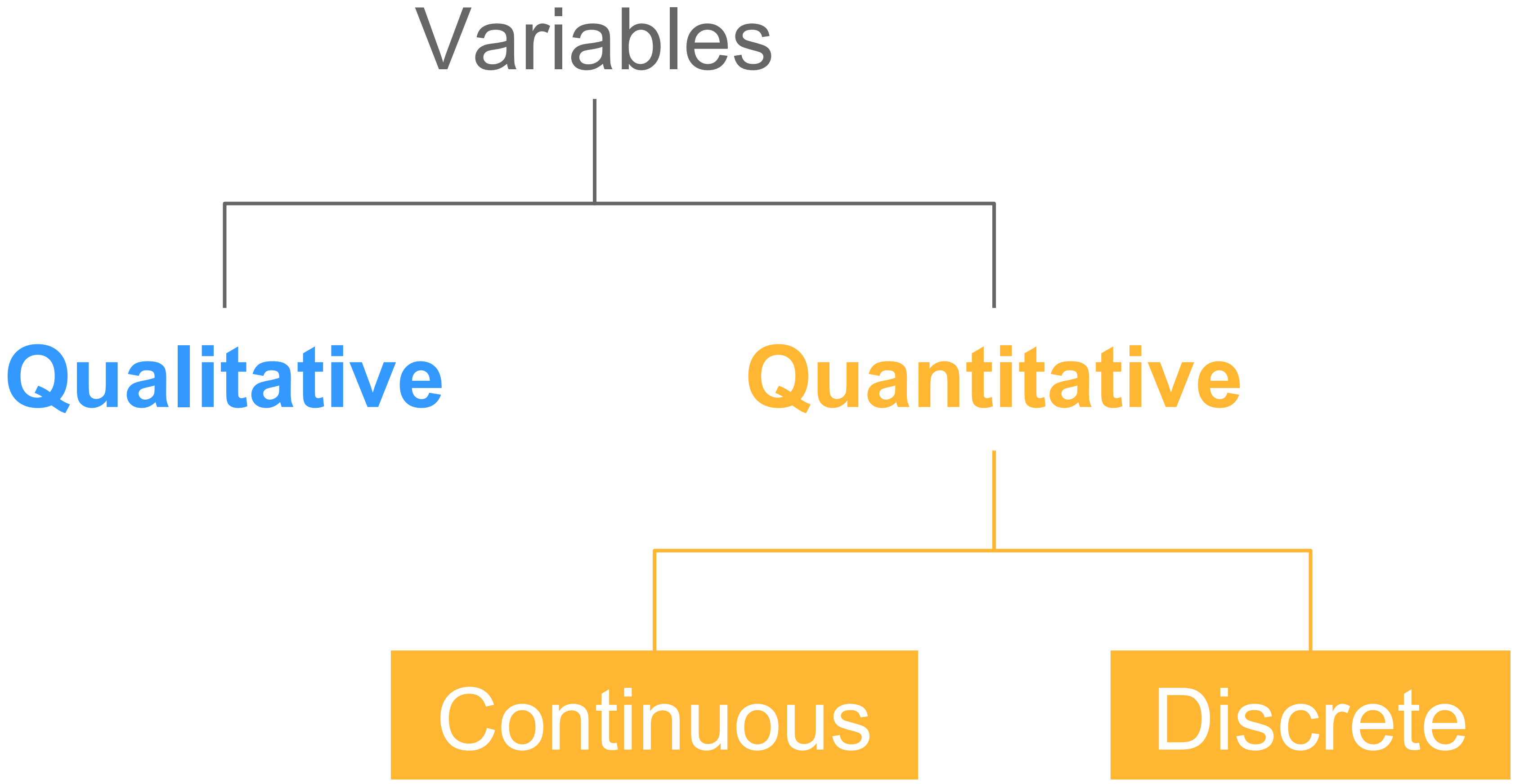
Quantifying

sugar content

milk content

pH (power of hydrogen)

More about quantitative variables



Variables

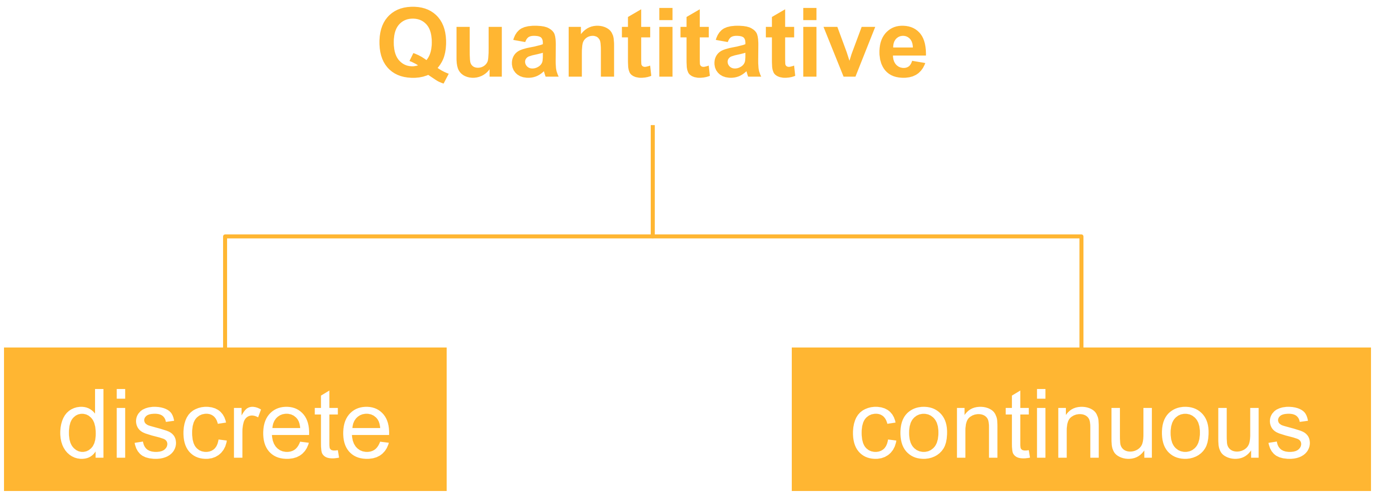
Qualitative

Quantitative

Continuous

Discrete

Quantitative Variables can also be divided in



Quantitative

discrete

continuous

Discrete Quantitative Variable

Takes on only a finite number of values or a countable number of values

Discrete Quantitative Variable

Number of days in a year

Number of laps you can swim in 5 mins

Number of touchdowns in superbowl

Continuous Quantitative Variable

Takes on any of the countless number of values in a line interval

Continuous Quantitative Variable

Time to run 100 meters

Distance while running 30 minutes

Size of a text file