Descriptive Statistics: Barcharts

Gaston Sanchez

NBA season 2015-2016

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

Data file "nba_players.csv" available in the course's github repository

Some questions

What is the typical salary of an NBA player?

How many players earn this money?

Do Point Guards earn more than Centers?

Do players with higher salaries score more points?

Descriptive Statistics

What is the typical salary of an NBA player?

Measure of center

How many players earn this money?

Distribution

Do Point Guards earn more than Centers?

Variability

Do players with higher salaries score more points?

Association

Some questions

To measure player richness, do I have to consider everyone?

If I don't consider everyone, how reliable is my estimate?

How different is the salary between Point Guards and Centers? Is the difference really important?

Inferential Statistics

To measure player richness, do I have to consider everyone?

Sampling

If I don't consider everyone, how reliable is my estimate?

Confidence

How different is the salary between Point Guards and Centers? Is the difference really important?

Hypothesis test

Describing Data

1 Frequency Tables
2 Charts & Graphics
3 Numeric Summaries

Main Idea:

Make a large or complicated dataset more compact and easier to understand by organizing it in a table, chart, or graph.

Example with a qualitative variable

Player Position

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

Visualization with a bar-chart

What do we need to make a bar-chart?

A Frequency Table

Frequency table

Position	Count	Percent

Frequency table

Position	Count	Percent
Center		
Point Guard		
Power Forward		
Shooting Guard		
Small Forward		
total		

Frequency table

Position	Count	Percent
Center	96	
Point Guard	112	
Power Forward	112	
Shooting Guard	107	
Small Forward	101	
total	*528*	

Frequency table

Position	Count	Percent
Center	96	18.18%
Point Guard	112	21.21%
Power Forward	112	21.21%
Shooting Guard	107	20.26%
Small Forward	101	19.12%
total	*528*	*100%*

Making a frequency table

Frequency tables ...

Category Name	Frequency (counts)	Percent (proportion)
C_1	f_1	f_1 / n
C_2	f_2	f_2 / n
C_3	f_3	f_3 / n
...	...	...
C_k	f_k	f_k / n

f_i: # objects in *i*-th category C_i
n: total number of objects

Making a Frequency Table

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 Count the number of occurrences for each distinct value and record the totals in the second column of the table.

Step 3 Optionally, divide the counts by the total number of values to get the frequencies as percents.
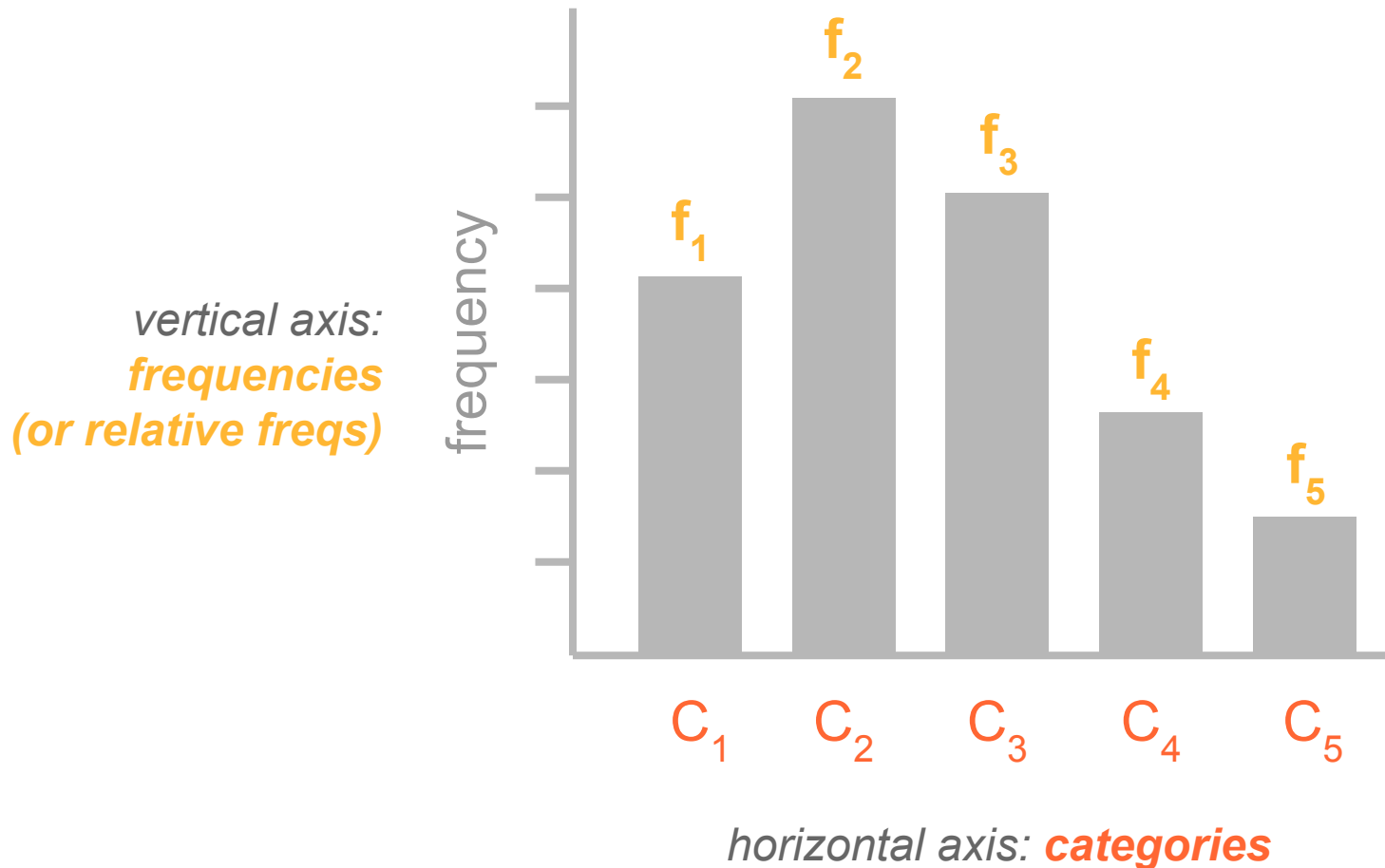
Making a barchart

Start with a frequency table

Category Name	Absolute Frequency	Relative Frequency
C_1	f_1	f_1 / n
C_2	f_2	f_2 / n
C_3	f_3	f_3 / n
...	...	...
C_k	f_k	f_k / n

Elements of a Bar-chart *(vertical bars)*

vertical axis:
**frequencies
(or relative freqs)**

frequency

f_2

f_3

f_1

f_4

f_5

C_1 C_2 C_3 C_4 C_5

horizontal axis: **categories**

Main characteristics of bar-charts

Bars can be rearranged

Can be plotted vertically or horizontally

Bars don't need to be adjacent

What matters is the length of the bar (not thw width)