

STAT 1350, 2/12 Discussion Questions

1. A company database contains the following information about each employee: age, date hired, sex (male or female), ethnic group (Asian, black, Hispanic, etc.), job category (clerical, management, technical, etc.), and yearly salary. Which of the variables are categorical?

Circled

2. Is a person's test score on the SAT a good predictor of that person's future college grade point average (GPA)? A researcher gathers data on the SAT score and college GPA for 569 college seniors. These measurements are examples of quantitative or categorical data?

quantitative

SAT scores are #'s & so is GPA

3. You have data on returns on common stocks for all years since 1945. To show clearly how returns have changed over time, what is your best choice of graph?

line graph

4. A bar graph compares the size of the armed forces for China, North Korea, Russia, and the United States. To make the graph look nicer, the artist replaces each bar by a proportionally correct picture of a soldier that is enlarged or reduced to be as tall as the bar. Why is this graph considered misleading?

because pictures scale by area not just height

5. Describe seasonal variation, and give three examples of it.

cyclic patterns in data : yearly rainfall ; daylight hours, store sales , etc.

Does using a cell phone while driving make an accident more likely? Researchers compared telephone company and police records to find 699 people who had cell phones and were also involved in an auto accident. Using billing records, they compared cell phone use in the period of the accident with cell phone use the same period on a previous day. Result: The risk of an accident was four times higher when using a cell phone.

6. The researchers also recorded the manufacturer of each subject's cell phone (Apple, Samsung, etc.). This variable is categorical or quantitative?

Categorical

7. The proper graph for showing the distribution of phones by manufacturer (i.e., number of people who own an Apple phone, number of people who own a Samsung phone, etc.) is what kind of graph? Why?

pie chart (represents parts of a whole)
bar graph is ok.

8. When pictures replace the bars in a bar graph, the resulting graph is called what?

pictograph

9. The proper graph for showing the percentage of students in a monogamous relationship, grouped by year in school (freshman, sophomore, etc.) is what?

bar graph (not part of a whole)

10. In order to create a good graph, what are three things you should do?

labels & legends
make data stand out
what does the eye see?

11. What tells us what values a variable takes and how often it takes those values?

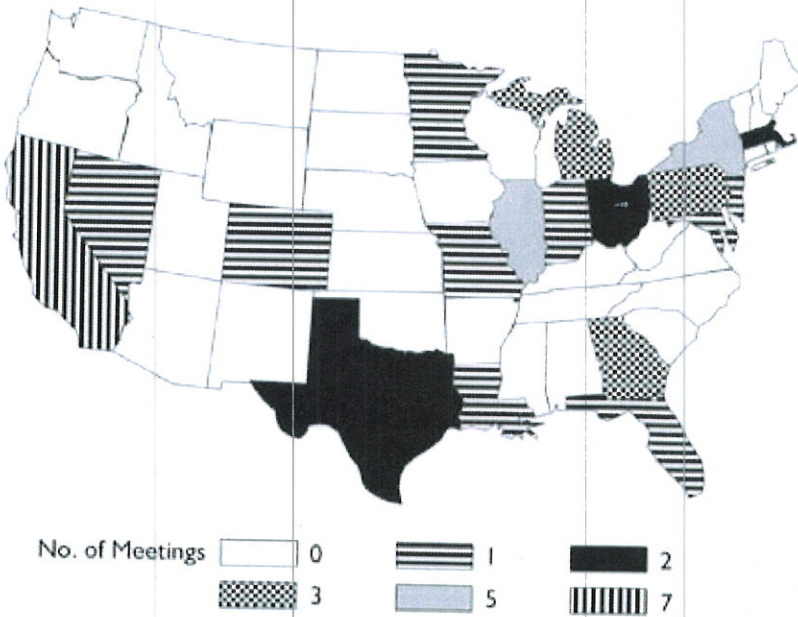
distribution

12. What does "seasonal adjustment" mean?

expected seasonal variation is removed

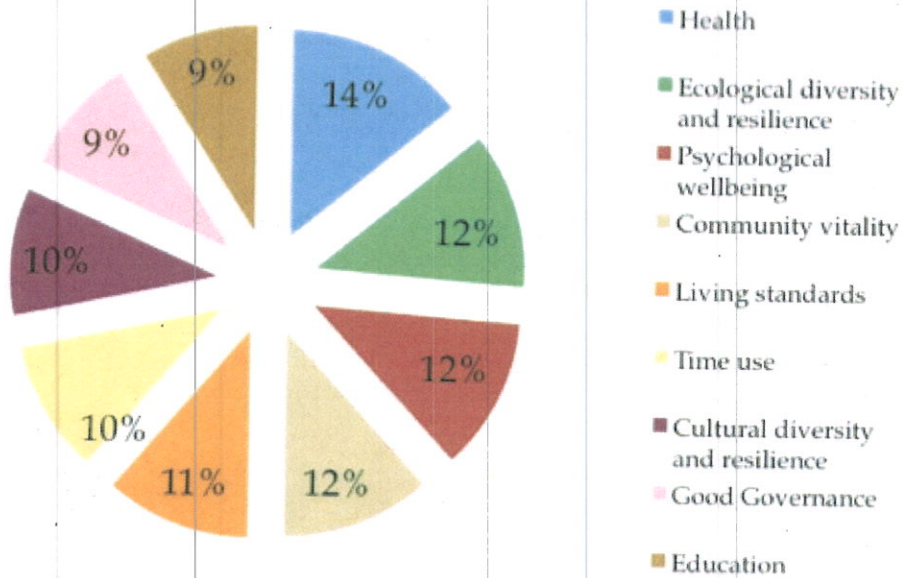
STAT 1350, 2/17 Discussion Questions

Below are a series of graphs, explain why they are good or bad graphs.

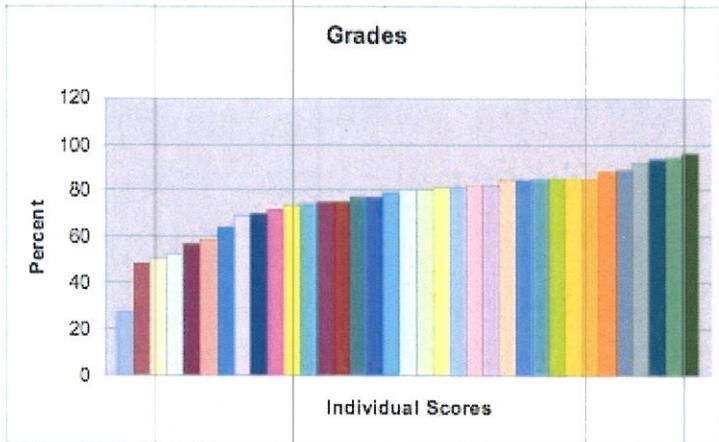


*bad.
difficult to look at
unclear what is
being measured.*

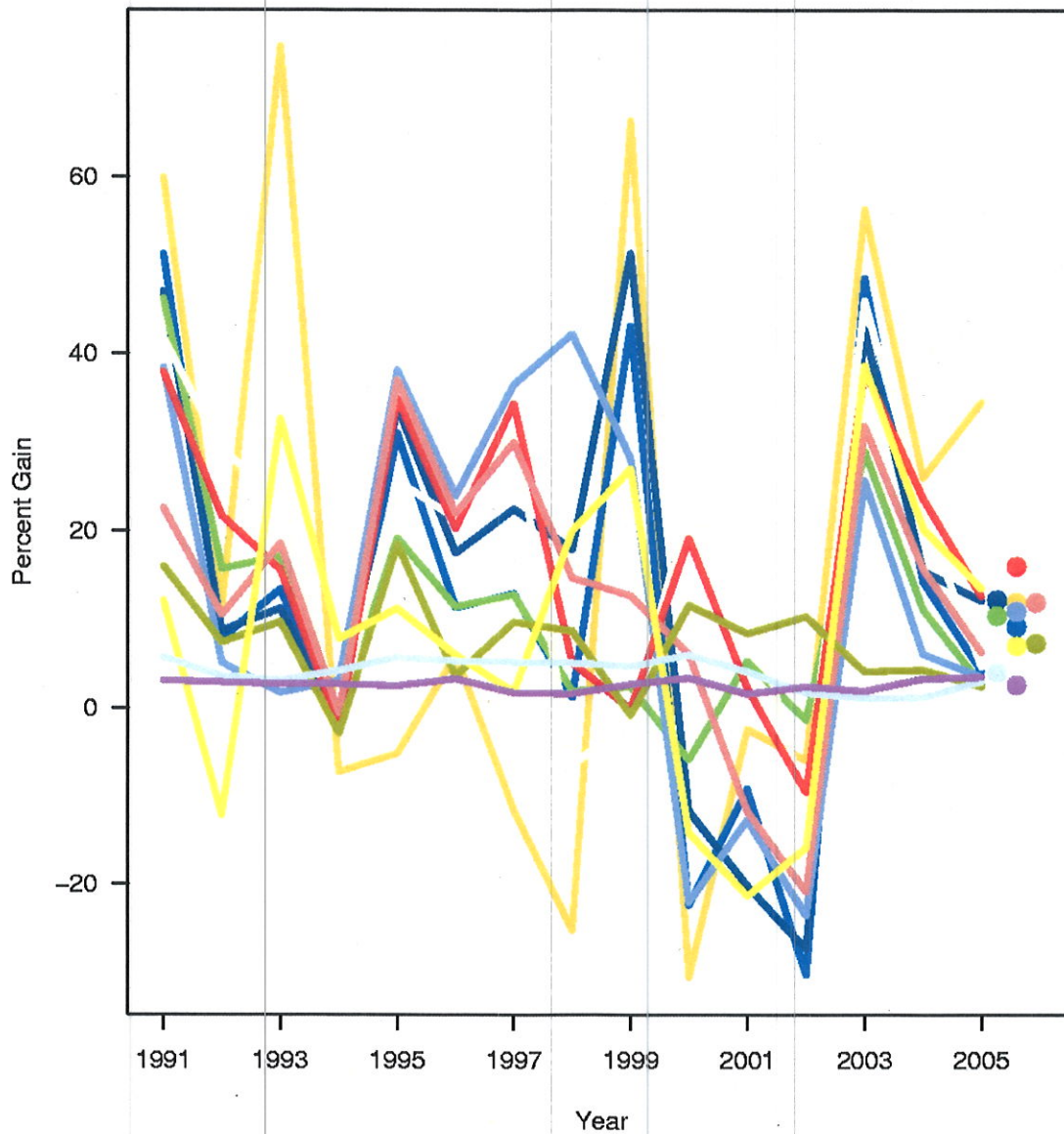
Figure 4: In which domains do happy people enjoy sufficiency?



*if respondents can only choose 1, this may be okay
(99% could be a rounding error), however,
the wording of the question does not imply this fact
so a bar graph is better*



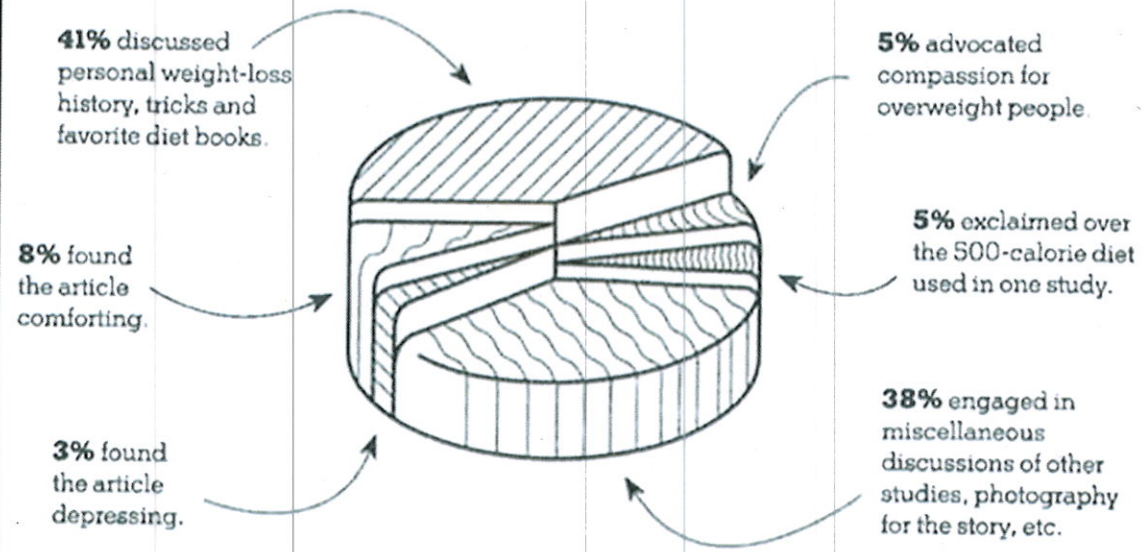
bad.
 difficult to read.
 taller bars appear to
 have more weight.
 Should use histogram
 for quantitative data
 like test scores



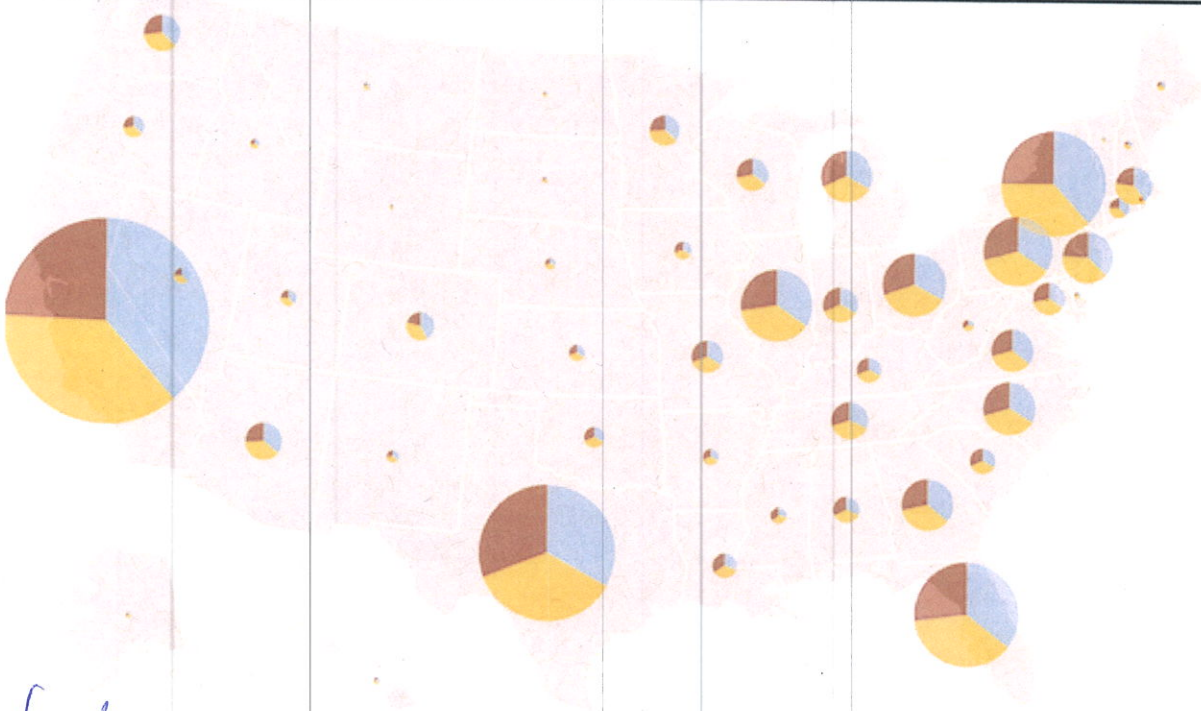
bad. too busy. hard to read
 too many lines are too similar in color

© ANALYTICS: A Steady Diet of Comments

Of 1,047 comments on Tara Parker-Pope's article, "The Fat Trap":

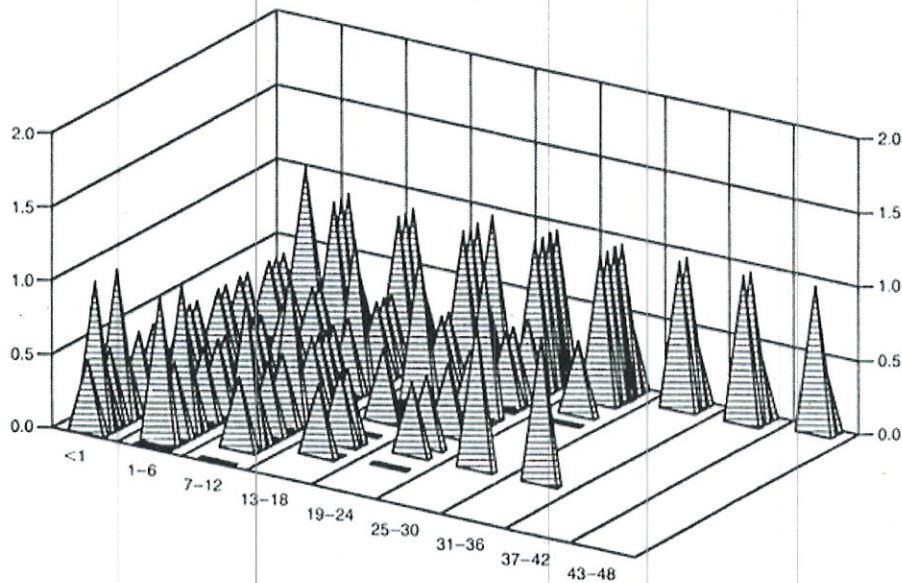


bad
don't use
perspective
in a pie
chart:
near piece
look bigger
than far
ones.
volume
makes
this worse.



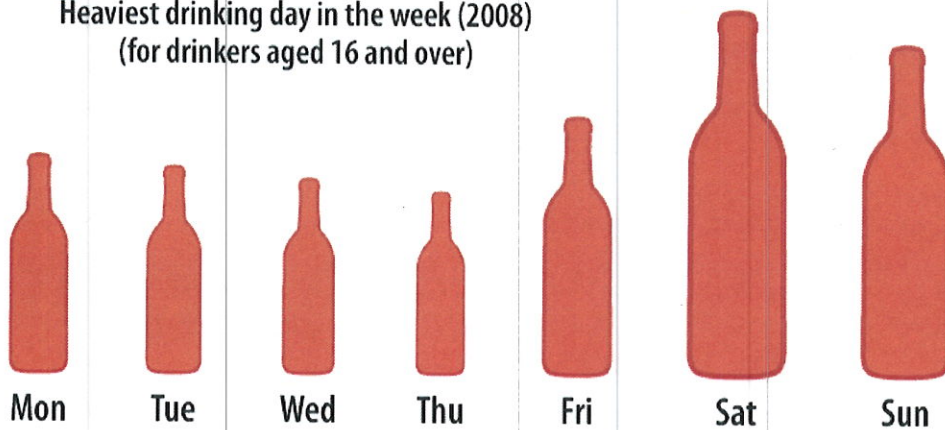
bad

pie charts for each state are terrible.
many are too small to read, some overlap
difficult to compare
no label or legend.

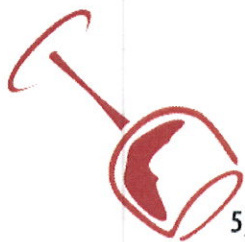


bad, especially in 2D. Small bars hidden behind taller ones might work if you can rotate it.

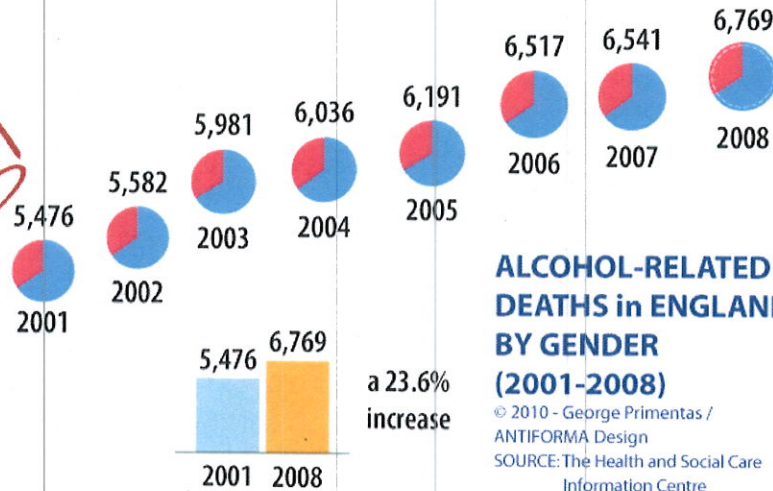
Heaviest drinking day in the week (2008)
(for drinkers aged 16 and over)



bad. Scaled by area when they intend only height noscale



35% ♀
65% ♂
GENDER OF THE VICTIMS

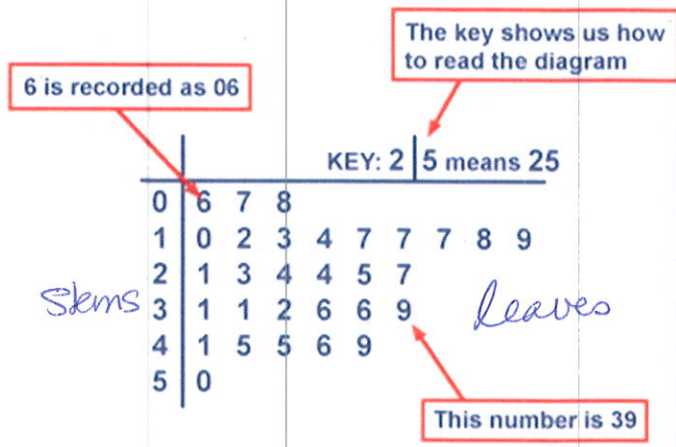


ALCOHOL-RELATED DEATHS in ENGLAND, BY GENDER (2001-2008)

© 2010 - George Primentas / ANTIFORMA Design
SOURCE: The Health and Social Care Information Centre

bad. 2 line graphs would be better (or 3) difficult to compare a time series of pie charts.

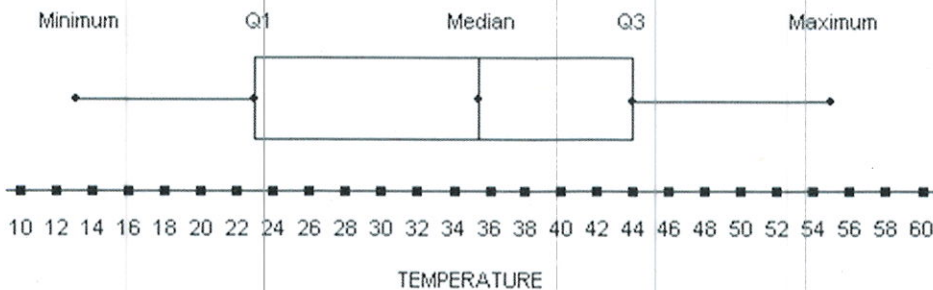
For each of the graphs below, describe what kind of graph it is and what kind of data (categorical or quantitative) it is displaying. Label all key features of the graph. What are the steps necessary to create each type of graph?



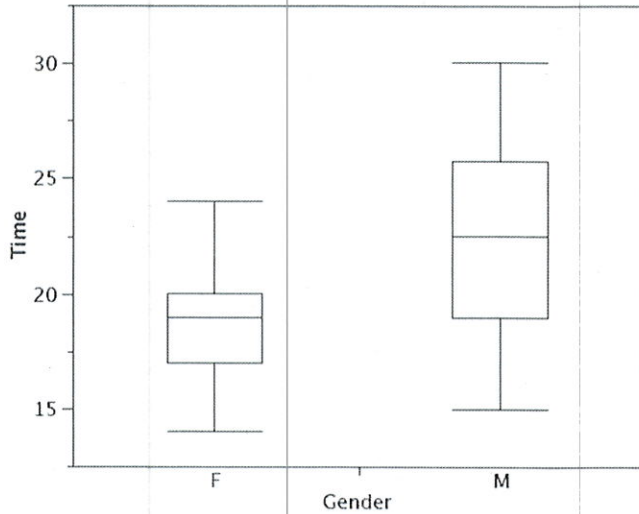
- Stem plot
quantitative
- Sort data
 - choose stem - more than 5, not too many
 - put down leaves
 - create key

Male		Female
5, 2, 0	1	5, 8
5, 1	2	1, 6, 9, 9
5, 5, 5, 3, 1	3	
5, 2	4	1, 2, 6, 8
9, 8, 6, 1, 1	5	5
6, 5, 5, 0	6	0, 1
2, 1, 1, 0, 0	7	2

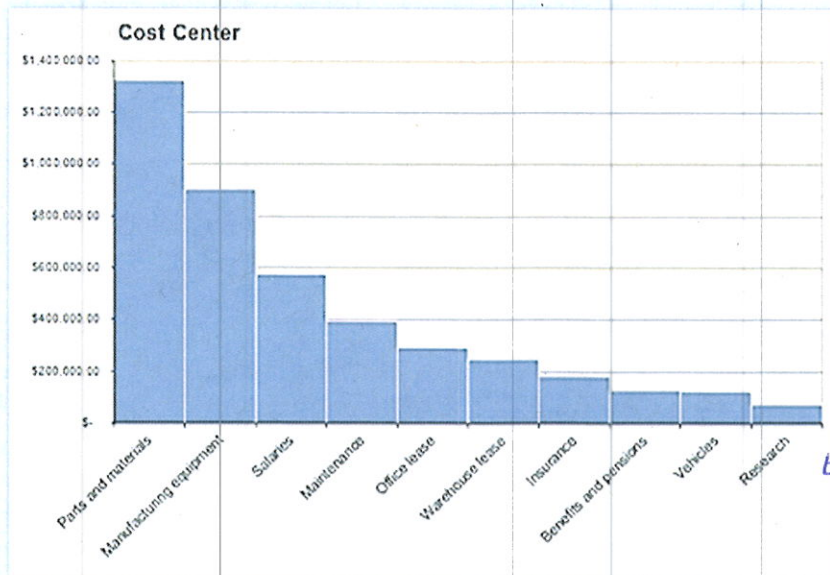
- Comparative
Stem plot
quantitative
- (steps above)
repeat for 2nd set of data



- boxplot
quantitative
- Sort data.
 - find min & max
 - find median
 - find Q₁ & Q₃
 - draw scale.
 - box between Q₁, Med, Q₃. whiskers to min, max.

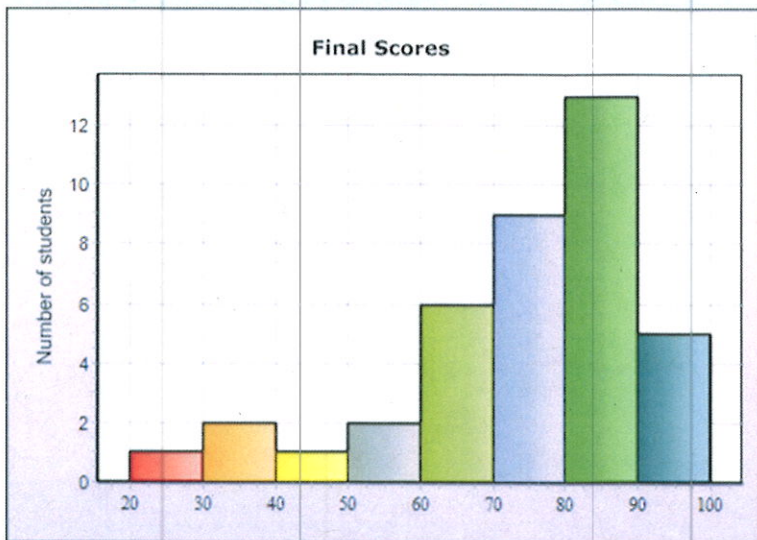


(Comparative) box plot
quantitative
(see above)



bar chart /
pareto chart

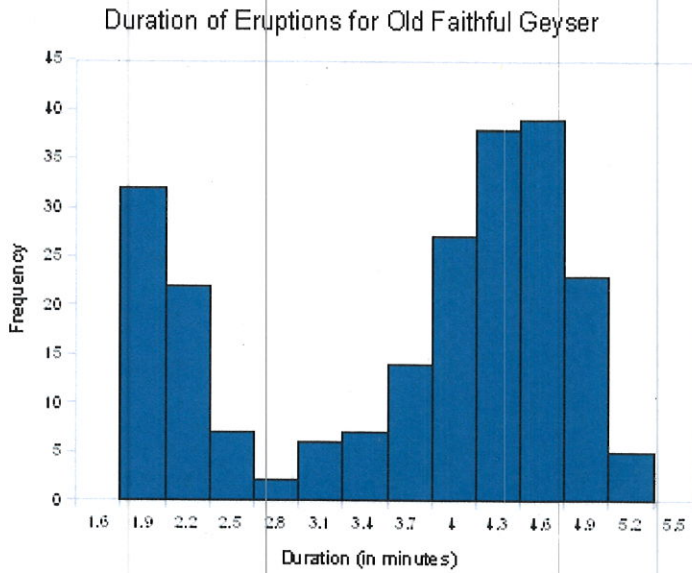
← categorical data



bar chart / histogram

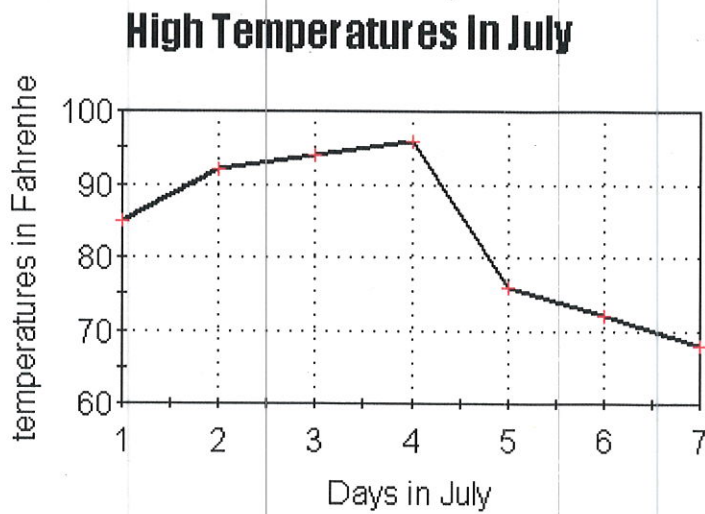
← quantitative data

- ① group data into categories
- ② plot bars to show # of people in each group.

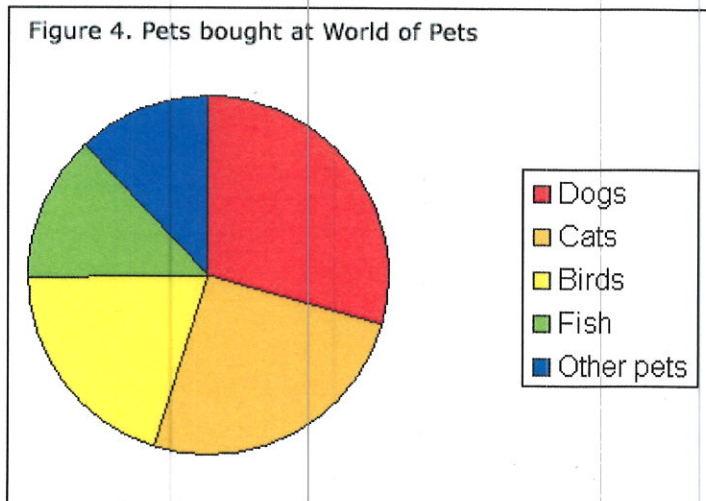


histogram
(see above)

← quantitative



line graph
quantitative
① plot points. ② connect dots.



pie chart
categorical

① calculate proportions for each category
② convert to angles by $\times 360^\circ$.

STAT 1350, 2/19 Discussion Questions

1. Here is a set of data: 1300, 18, 25, 19, -7, 24. Which observation is the outlier?

1300, (-7 maybe)

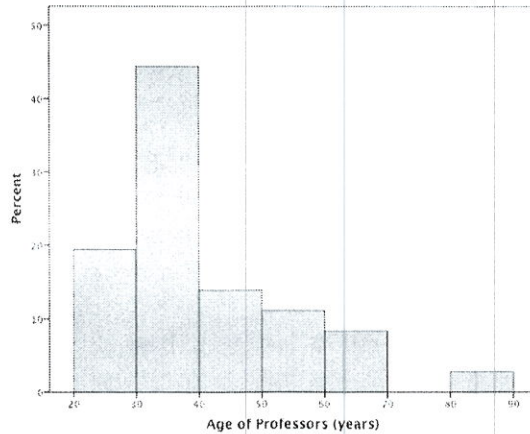
2. To display the number of pets owned by each of the 37 students in a class, what would be a good choice of graph?

bar chart

3. You want to make a graph that shows how the cost of attending your school has increased since 1980. What would be a good choice of graph?

line graph

Below is a histogram of the ages of the professors at a large university.



4. The overall shape of this distribution is generally symmetric, skewed right or skewed left?

skewed right (tail)

5. Approximately what percentage of the professors are 30–39 years old?

45%

6. There are 800 professors at this university. How many of them are in their 70s?

possibly zero

(fewer than 8)
1%