Liberian Mathematics Teacher Training Workshop 2023–2024

Andrew Obus[1]

CUNY Baruch College

September 22, 2023

[1]This program is partially supported by NSF CAREER Grant DMS-2047638

Graphical presentation of data

- Pie charts
- (Comparative) bar graphs
- Stem and leaf plots
- Histograms
- Cumulative frequency curves → Next time
- Box and whisker plots → after discussing measures of central tendency and dispersion

Categorical data

- *Categorical data* is data collected that does not consist of numbers. For instance, a survey might ask a group of people whether they are smokers or non-smokers, or what political party they belong to. Each possible response is called a *category*.
- So, for example, in the United States, the categories for the political party question might be “Democratic”, “Republican”, “Other”, or “Independent”.
- In Liberia, there would probably be many more categories!
- Categorical data, especially when the number of categories is relatively small, can often be effectively presented as a *bar graph* or a *pie chart*.

Categorical data: pie charts

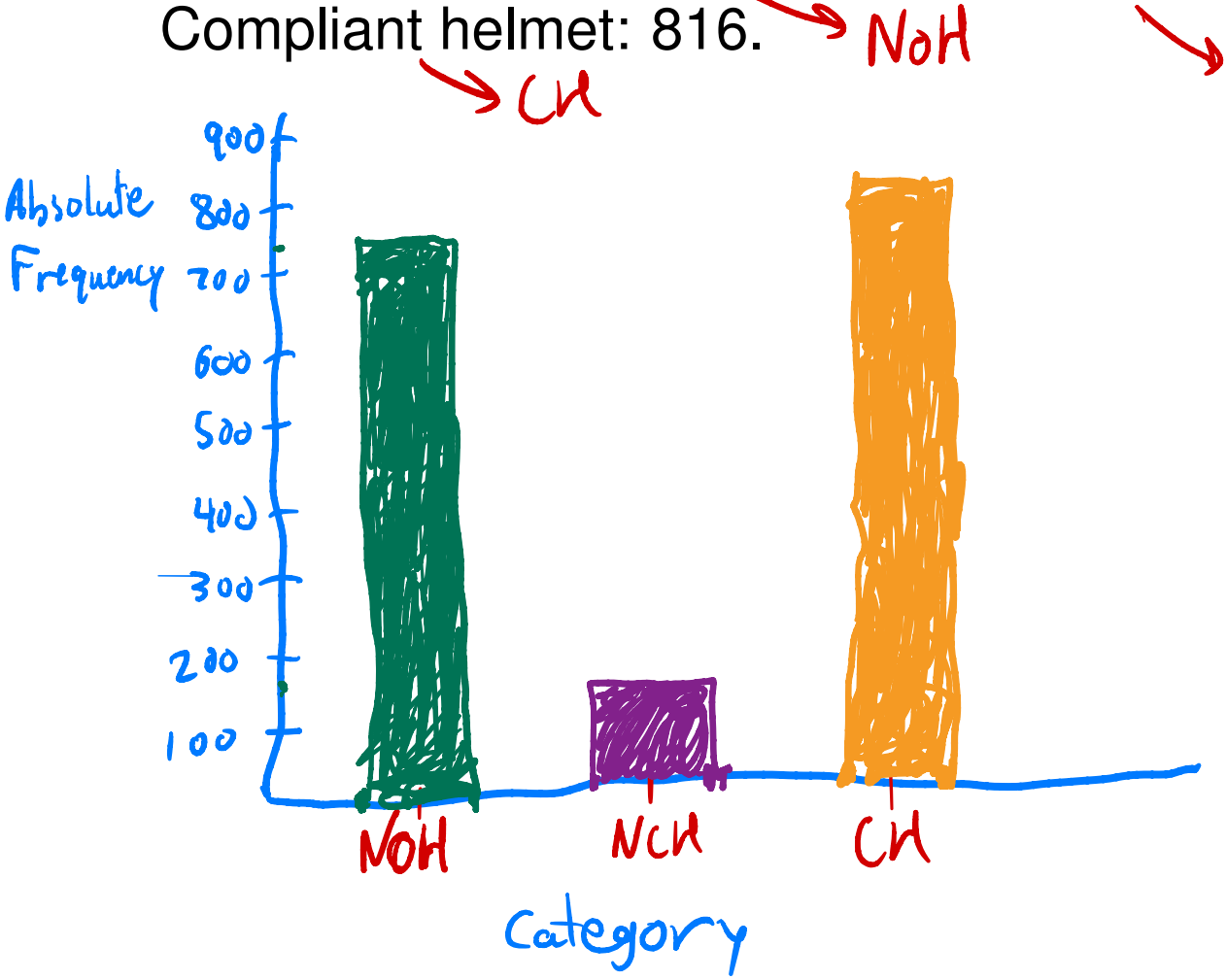
- Liberia ethnic group data (2008): Kpelle 20.3%, Bassa 13.4%, Grebo 10.0%, Gio 8.0%, Mano 7.9%, Kru 6.0%, Loma 5.1%, Kissi 4.8%, Other 24.5%
- Constructing by hand is often quite difficult. Usually done with software such as Microsoft Excel, Google Sheets, or OpenOffice Calc.
- Can also do on https://nces.ed.gov/nceskids/graphing/classic/pie.asp
- That is what we will do.

Categorical data: bar graphs (charts)

- Constructing bar graphs by hand is a bit easier.
- On the horizontal axis, we list the categories, evenly spaced.
- The vertical axis represents frequencies — absolute, or relative (as percentage).
- Can also do on all of the software packages mentioned above, as well as on https://nces.ed.gov/nceskids/graphing/classic/pie.asp.

Bar Graph: Example

- Motorcycle helmet use in USA, 2005. 1700 motorcyclists observed. No helmet: 731. Noncompliant helmet (too small): 153. Compliant helmet: 816.



Frequency Table

Category	Frequency
NoH	731
NCH	153
CH	816
Total	1700

Relative Frequency Table

Category	Relative Frequency
NoH	.43
NCH	.09
CH	.48

$\frac{731}{1700} = .43$ $\frac{153}{1700} = .09$ $\frac{816}{1700} = .48$

$$\text{Relative Frequency} = \frac{\text{Absolute Frequency}}{\text{Total \# of responses.}}$$

Bar Graphs vs. Pie Charts

- It is not always clear which is better to use, a pie chart or a bar graph.
- If there are lots of categories, often a bar graph is better (pie charts can get cluttered and difficult to read).
- If the important concept is what proportion of the whole each category takes up, then a pie chart can make more sense.
- If there is no real concept of "the whole", then a bar graph is better.
  - Share of Liberia's land with each usage (agricultural, forest, settlement)? → Pie Chart (bar graph also ok)
  - ~~Favorability rating of a politician over 12 different months of 2022?~~

Come up with an example where a bar graph makes more sense than a pie chart.

Numerical data → "Quantitative Data"

- Stem and leaf plots
- Histograms from frequency tables
- Cumulative frequency curves (NOT TODAY)
- Box and whisker plots (NOT TODAY)

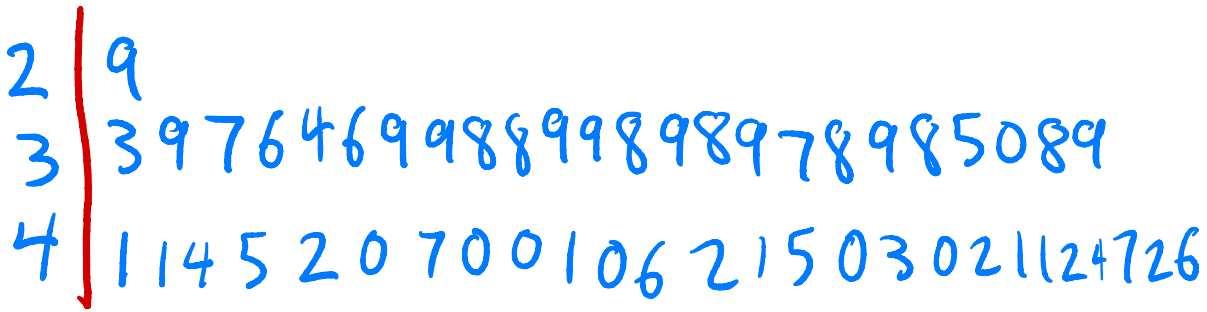
Stem and leaf plots

- Useful when you have numerical data with not too many observations.
- Gives a direct way for turning numerical data into a suggestive picture.
- Idea is that we take some number of leading digits in the data. These are the stem values.
- Then, we take some number of trailing digits to be the leaves.
- You have to be **consistent** within your data set about where you split the stem from the leaf (always at the same place value). Sometimes this will result in the stems for certain entries having different numbers of digits or being “0”.
- An example will make this clearer.

Stem and leaf plots, example

Median ages in 50 US states in 2030 (projected): 41 33 39 29 37 36 41
44 34 45 36 39 39 38 38 42 39 40 39 47 38 40 40 39 41 40 46 38 39
42 41 45 40 37 43 40 38 39 42 41 41 42 38 35 30 44 38 39 47 42 46

- In this case, it makes sense just to use the first digit to be the stem and the second digit to be the leaf.
- Can do by hand. Can also use a website such as https://www.calculatorsoup.com/calculators/statistics/stemleaf.php (this has the benefit of automatically sorting the leaves)
- Can also do what are called “split stems”, where each stem is listed twice, once for higher leaf values (56789), and once for lower leaf values (01234).



Stem and leaf plots, second example

Percentage of primary-school age children enrolled in school by country in Northern Africa:

- 54.6 34.3 48.9 77.8 59.6 88.5 97.4 92.5 83.9 96.9 88.9 98.8 91.6 97.8 96.1 92.2 94.9 98.6 86.6
- If we use the first digit (the tens digit) for the stem, we have a choice of using just the ones digit for the leaf, or both the ones digit and the decimal place. We will do a plot for each choice.

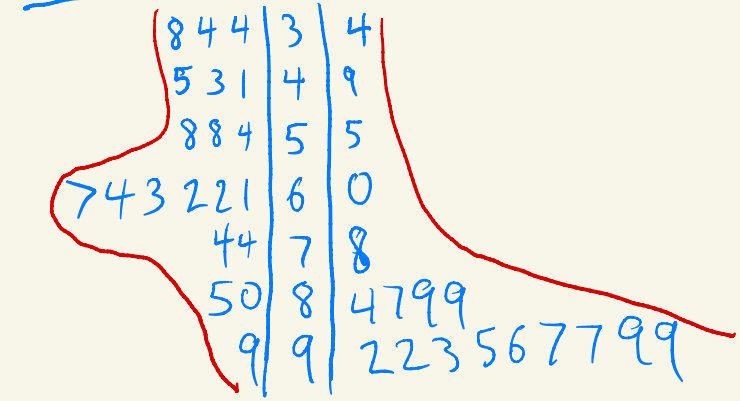
Comparative stem and leaf plots, second example

Stem and leaf plots can be used to compare *two* sets of data. For instance: percentage of primary-school age children enrolled in school by country

- NORTHERN AFRICA: 54.6 34.3 48.9 77.8 59.6 88.5 97.4 92.5 83.9 96.9 88.9 98.8 91.6 97.8 96.1 92.2 94.9 98.6 86.6
- CENTRAL AFRICA: 58.3 34.6 35.5 45.4 38.6 63.8 53.9 61.9 69.9 43.0 85.0 63.4 58.4 61.9 40.9 73.9 34.8 74.4 97.4 61.0 66.7 79.6 98.9

- In this case, the same data is collected from two different samples, so we have two data sets. In one set, we expand the leaf to the left, and in the other set, we expand the leaf to the right.
- To simplify the presentation of this data, we will use only one digit for the stem and one digit for the leaf.

Comparative Stem and leaf plot



% of primary school age children enrolled in school by country

7 3 7 9

2 8 6 2 5 9

123, 125, 246, 257, 300

1	23, 25
2	46, 57
3	00

OR

1	22
2	45
3	0

(tens)

Homework exercises

1. Suppose a survey is given where people are asked about the highest level of education they have completed: Primary School, Secondary School, University, or Graduate School. What do you think would be the most appropriate way to visually present the resulting data?
2. Make a stem and leaf plot for the following data: Liberia: Population by county (2008): Bomi: 82036, Bong: 328919, Gbarpolu: 83758, Grand Bassa: 224839, Grand Cape Mount: 129055, Grand Gedeh: 126146, Grand Kru: 57106, Lofa: 270114, Margibi: 199689, Maryland: 136404, Montserrado: 1144806, Nimba: 468088, Rivercess: 65862, River Gee: 67318, Sinoe: 104932. Try to choose the stem and the leaves so that the plot you get is meaningful!

Thank you for your attention. Next week we will talk more about frequency tables and classes, and also discuss cumulative frequency plots.

↳ Histograms