

Course Information

- [Schedule and Homework](#)
- [Course Requirements](#)
- [Instructor's Information](#)
- [Links](#)
- [Main Page](#)

Topics

1. [Introduction](#)
2. [Summarizing Data: listing and grouping](#)
3. [Summarizing Data: Statistical Descriptions](#)
4. [Possibilities and Probabilities](#)
5. [Some rules of probability](#)
6. [Probability Distributions](#)
7. [Normal Distributions](#)
8. [Sampling Distributions](#)
9. [Problems of Estimation](#)
10. [Hypothesis Tests](#)
11. [Tests based on count data](#)
12. [Regression](#)
13. [Non-parametric Tests](#)

Projects

1. [Stock return distribution](#)

- [Forum](#)

[edit](#) [sec](#) [src](#) [prt](#)

[Site Manager](#)

Chapter 2. Summarizing Data: listing and grouping

[Fold](#)

Table of Contents

- [Introduction](#)
- [Tables](#)
- [Frequency distribution table](#)
- [Stem-and-Leaf Plots](#)
- [Graphical representations](#)
- [Histogram](#)
- [Frequency Polygon](#)
- [Cummulative frequency distributions and Ojives](#)
- [Dot plot](#)
- [Pie Chart](#)

Introduction

Descriptive Statistics describe basic features of the data gathered from an experimental study in various ways. They provide simple summaries about the sample via graphs and numbers, mainly measures of center and variation.

Together with graphics analysis (histograms, bar plots, pie-charts), they are the cornerstone of quantitative data analysis.

- Tables (frequency distributions, stem-and-leaf plots, ...) that summarize the data.
- Graphical representations of the data (histograms, bar plots, pie-charts).
- Summary statistics (numbers) which summarize the data.

Tables

The most common ways of summarizing data into tables are frequency distribution, relative frequency distribution and relative frequency distribution tables. Another common format is using a stem-and-leaf plot.

Frequency distribution table

A **frequency distribution** summarizes the data into a table containing ranges where the data falls, and the frequency (or amount) of data that fall in that range.

For example, the frequency distribution of the [GDP per capita](#) in the 20 countries with the highest such gross domestic product per person is given below:

| GDP range | Number of countries |
|---------------|---------------------|
| - 40,000 | 5 |
| 40,000-49,999 | 8 |

| | |
|---------------|---|
| 50,000-59,999 | 4 |
| 60,000-69,999 | 0 |
| 70,000-79,999 | 2 |
| 80,000 + | 1 |

This means that among those countries, 5 have a GDP per capita of 40,000 dollars or less, 8 have a GDP per capita of between 40,000 and 49,999 dollars, and so on and so on.

Notice that such a table is already a summary, since we do not know exactly what the GDP of each country is, we only know that it is in the given range.

Another way of summarizing data in which we loose less information is using a stem-and-leaf plot.

Stem-and-Leaf Plots

Another way of summarizing data into tables is the stem-and-leaf plot, which works best when the data can be subdivided into tens and units. Suppose that the GDP per capita from the example above is:
81 76 70 57 55 55 89 46 46 45 44 44 42 40 39 39 56 39 39 39, which is given in thousands of dollars. In other words, one country has a GDP per capita of 81,000 dollars (namely Luxemburg), Qatar has a GDP per capita of 76,000, and for example the USA have a GDP per capita of 46,000 dollars. This data can be nicely summarized in the following stem-and-leaf-plot:

The decimal point is 1 digit(s) to the right of the 1

```
3 | 99999
4 | 02445669
5 | 5567
6 |
7 | 06
8 | 1
```

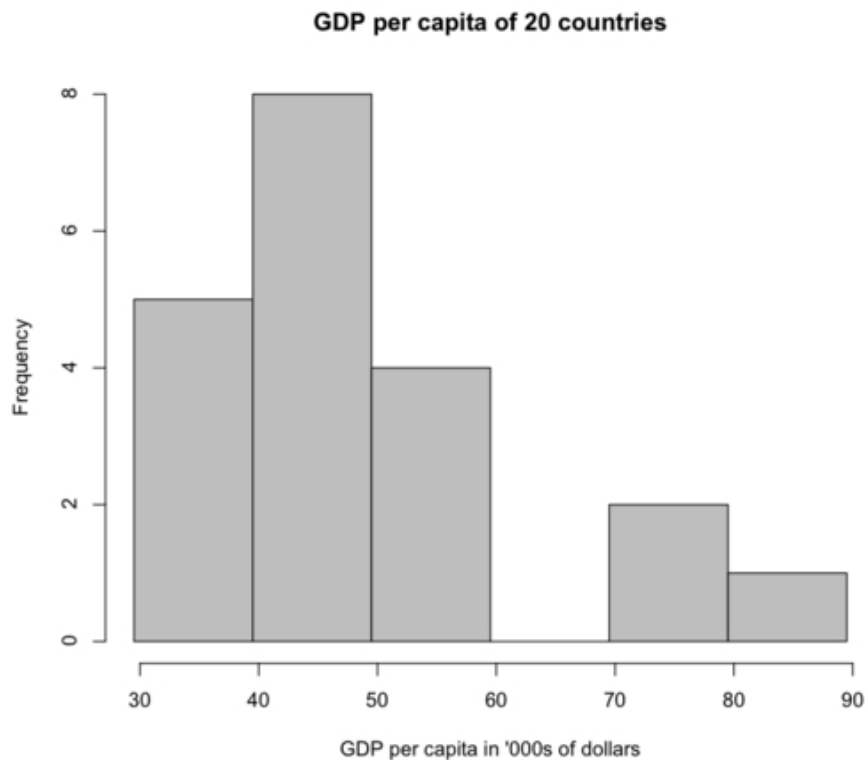
In the stem-and-leaf plot above, **8|1** means that there is one county with GDP per person in the 80,000s, and that that county has a GDP of 81,000 per capita.
7|06 means that there are 2 countries with GDP between \$70,000 and \$79,999, and that those countries have GDP per capita of **70,000** and **76,000**. That is why there is a 0 and a 6 in the units are, to the right of the vertical bar.

Graphical representations

Examples of graphical representations are the histogram, bar plot, dot plot and pie chart.

Histogram

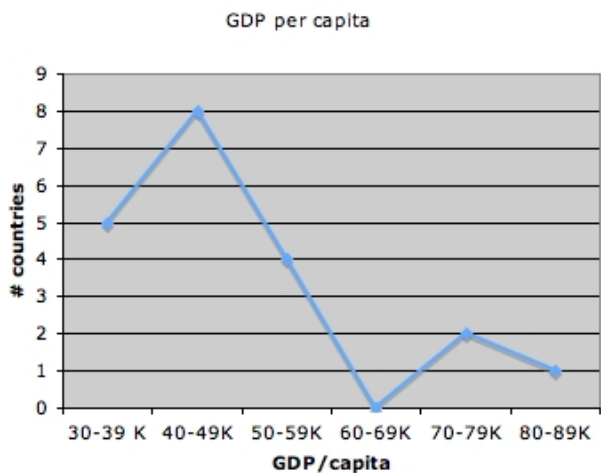
A histogram is a vertical bar chart, where the values heights represent the frequencies (also called counts). For the example about the GDP per capita, for which we made a frequency distribution above, the histogram would look like this:



Note that there is no space in between the bars, and that the height of the bars represent the frequencies, while the width of the bars depends on the class range selected in the frequency distribution.

Frequency Polygon

A frequency poligon is another way of graphically representing the data, where instead of using bars, we connect the given heights by a straight line, as in the example below.



Notice that this frequency polygon represents the same data given in the histogram above.

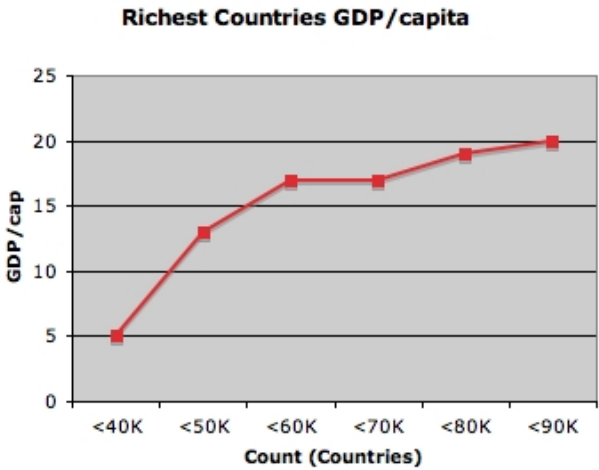
Cummulative frequency distributions and Ojives

A cummulative frequency table is obtained by counting how many items fall below a certain boundary. For example, from the frequency table above, we can create a new table, called a cummulative frequency table:

| GDP range | Number of countries |
|------------------|---------------------|
| less than 40,000 | 5 |

| | |
|------------------|----|
| less than 50,000 | 13 |
| less than 60,000 | 17 |
| less than 70,000 | 17 |
| less than 80,000 | 19 |
| less than 90,000 | 20 |

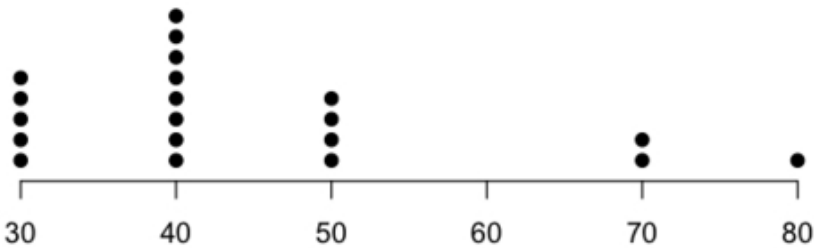
Which we can then use to create a cummulative frequency polygon, called an **ojive**.



Dot plot

There are several kinds of dot plots and dot charts. But we will concentrate on one that is similar to a histogram, but instead of bars, dots are piled on top of each other, as can be seen below:

Dot Plot of GDP per capita (in '000s of USD)



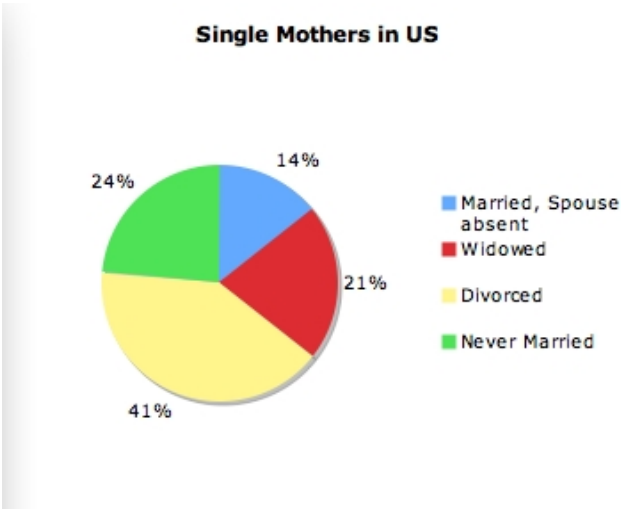
Pie Chart

A pie chart is a useful graphical representation for qualitative data.

For example, the [US Census Bureau](#) compiled the following chart for single mothers.

| Single Mothers: | | Thousands |
|------------------------|--|-----------|
| Married, Spouse absent | | 1,210 |
| Widowed | | 1,764 |
| Divorced | | 3,431 |
| Never Married | | 1,974 |
| Total | | 8,379 |

Using this qualitative chart, we can create a pie chart by dividing the value in every category by the total to convert to percent, and then multiplying by 360 degrees, to find the angle of the piece of the pie corresponding to that category.



<http://statistics.wikidot.com/local--files/ch2/Quiz1.xls>

page revision: 16, last edited: 1 Sep 2010, 18:06 (3057 days ago)

[Edit](#) [Tags](#) [History](#) [Files](#) [Print](#) [Site tools](#) [+ Options](#)

Powered by [Wikidot.com](http://www.wikidot.com) [Help](#) | [Terms of Service](#) | [Privacy](#) | [Report a bug](#) | [Flag as objectionable](#)

Unless otherwise stated, the content of this page is licensed under [Creative Commons Attribution-ShareAlike 3.0 License](#)