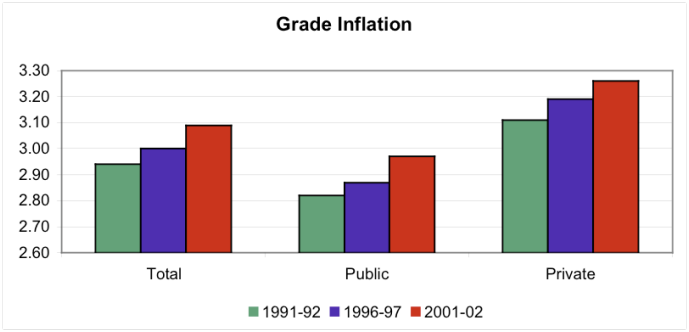
Exploring Data with Graphs

Dr Tom Ilvento
Department of Food and Resource Economics

Using graphs can be a useful way to tell a story with your data



Describing Data

- Statistics provides us tools to describe a set of data
- The tools involve both numerical and graphical summaries
- The first distinction is the nature of the data
  - Qualitative (Nominal and Ordinal)
  - Quantitative (Continuous)

Let's look at two small data sets

- Think about how you would summarize or describe the data using any tools you now have
- Think about how you might graph the data

Data Set 1

9.5	6.9	6.7	5.0	8.3
8.3	22.5	6.1	6.7	8.4
6.3	10.5	6.1	7.0	10.3
12.6	5.8	5.9	6.8	7.4
6.3	7.7	5.9	6.5	8.6
6.4	6.9	8.1	6.6	8.9
5.5	7.0	7.3	7.3	8.2
5.5	8.8	7.0	7.4	6.6
4.2	8.1	61.0	4.7	7.4
8.9	7.9	7.3	7.0	6.1
9.4				

Data Set 2

- 2 5 4 3 3 6 2 5 1 3 6
  2 2 3 4 5 3 3 2 2 4 3
  1 5 2 4 2 3 5 2 4 3 1
  1 3 2 6 2 3 3 2 2 3 1
  2 2 3 1 2 2

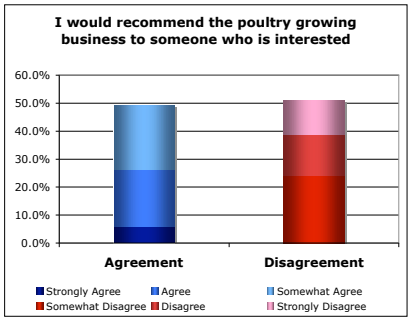
Data Set 1

- Data Set One was Marriage rate data for the U.S., 50 states and Washington D.C.
- It is the number of marriages in the state in a given year divided by the population, and then multiplied by 1000
- Quantitative data
- Leads itself to things like mean, mode, median (we'll discuss more about them in future lectures)
- And graphs such as Histogram, box plots, and stem and leaf plots

Data Set 2

- Data set 2 was the pattern of responses to a Likert scale of satisfaction with the poultry business from a survey of poultry growers on Delmarva
  - Strongly Agree, Agree, Somewhat Agree, Somewhat Disagree, Disagree, Strongly Disagree
  - Coded as 1, 2, 3, 4, 5, 6
- These data are Qualitative (or ordinal)
- Leads itself to frequencies, relative frequencies, and traditional graphing techniques

Poultry Grower Satisfaction



Qualitative Data Coding Schemes

- In qualitative data we had both nominal groups (no order or size implied) and ordinal (implied order or ranking)
- We often represent this data with numbers or categories
  - 0, 1 dichotomy can represent the presence of an attribute
  - 1, 2, 3, 4, 5…. Can represent categories or ordinal ratings

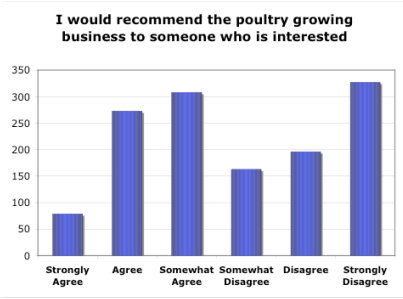
Qualitative Data Coding Schemes

- In qualitative data the categories are also referred to as **CLASSES**
- The number of observations in a class is called the **CLASS FREQUENCY**
- If we calculate a percentage for each class (of the total number of observations) we have the **CLASS RELATIVE FREQUENCY**
- And we can also calculate the **CUMULATIVE RELATIVE CLASS FREQUENCY**

I would recommend the poultry growing business to someone who is interested

Class	Code	Freq	Rel. Freq	Cum Rel. Freq
Strongly Agree	1	79	5.9%	5.9%
Agree	2	273	20.3%	26.2%
Somewhat Agree	3	308	22.9%	49.0%
Somewhat Disagree	4	163	12.1%	61.1%
Disagree	5	196	14.6%	75.7%
Strongly Disagree	6	327	24.3%	100.0%
TOTAL		**1346**		

And we can graph this is any of the following graphs



Look at the following data

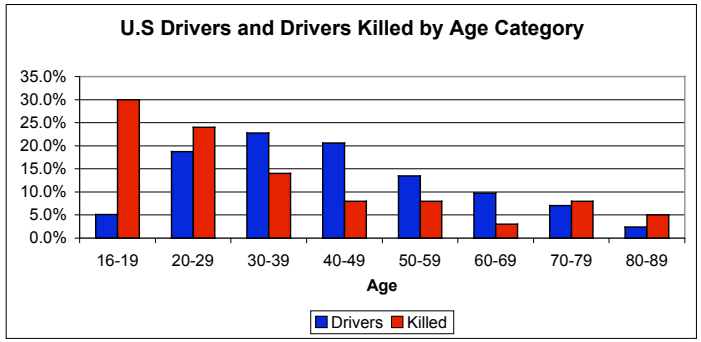
- This is a small data set of the age of 100 randomly sampled drivers killed in traffic accidents
- It is really continuous, but let's collapse the data into categories

16	17	18	20	24	28	34	44	53	73
16	17	18	20	24	28	35	44	56	74
16	17	18	20	24	29	36	45	57	76
16	17	18	21	24	29	37	45	58	76
16	17	18	21	25	30	37	45	65	77
16	17	18	21	26	30	38	45	69	81
16	18	18	22	27	31	38	51	69	86
16	18	18	23	27	31	38	51	70	87
17	18	18	24	27	32	40	51	73	87
17	18	19	24	28	34	42	52	73	88

Table Percentage by Age of a random sample of drivers killed in a car accident

Age Categories	Sample of Drivers Killed	Percent
16-19	30	30.0%
20-29	24	24.0%
30-39	14	14.0%
40-49	8	8.0%
50-59	8	8.0%
60-69	3	3.0%
70-79	8	8.0%
80-89	5	5.0%
TOTAL	100	

A Graph of the Table Data



Be careful when collapsing a continuous level variable - you almost always lose information and risk distorting the data.

A note on dummy variables

- One strategy for analysis of categorical variables, and sometimes ordinal variables, is to convert them to dummy variables.
- This is a strategy to convert the categories to a numerical value of either 1 or 0.
- Modern software has reduced the need to convert to dummy variables, but you should still be aware of this approach.

Dummy variables

- Any categorical variable with k categories can be represented with k-1 dummy variables
- A dummy variable takes on only two values, either zero or one. A value of one represents the presence of the attribute or category level.
- The category that is not represented by a dummy variable (the k-1 category) is referred to as the Reference Category and often serves as the comparison.
- When working with Dummy Variables, you must be clear on:
  - Which category is the reference category
  - What a one or zero represents

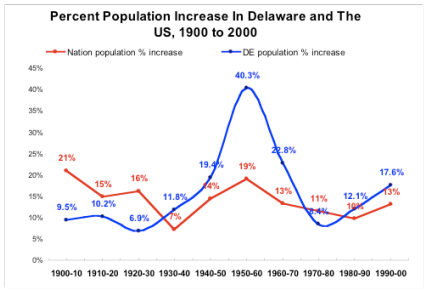
Dummy Variables

- **A survey question that asks for Yes or No response**
  - 2 categories, Yes or No. So k=2.
  - We use one dummy variable – choose whether Yes = 1 or No = 1.
- **Treatment Level of subjects in an experiment**
  - 3 categories – High Treatment, Low Treatment, No Treatment (control), so k = 3
  - We would create 2 dummy variables:
  - HIGH 1 = High Treatment, 0 = No
  - LOW 1 = Low Treatment, 0 = No
  - The control group would be the reference category

A few ideas on graphing qualitative data

- Many software programs can help you graph data
  - Excel or a spreadsheet
  - Word or word processing
  - Statistical software
- Excel provides many graphing strategies (14 types), with options on colors, layout, and design
- Excel will do:
  - Bar/Column
  - Pie
  - Area
  - Line
  - XY Scatter Plot

Excel graphing examples



Source: U.S. Census of Population

Checklist for Graphs

- Does the message of interest stand out clearly?
- Is the purpose or title of the picture evident?
- Is a source given for the data, either with the picture or in an accompanying article?
- Did the information in the picture come from a reliable, believable source?
- Is everything clearly labeled, leaving no ambiguity?
- Do the axes start at zero or not?

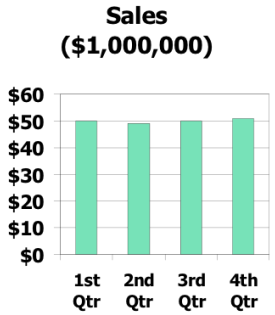
Checklist for Graphs

- Do the axes maintain a constant scale?
- Are there any breaks in the numbers on the axes that may be easy to miss?
- For financial data, have the numbers been adjusted for inflation?
- Is there information cluttering the picture or misleading the eye?

Seeing Through Statistics, Jessica M. Utts, 1999 Brooks/Cole Publishing Company

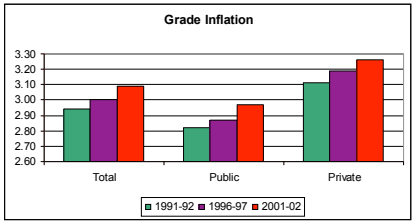
Some of my thoughts on bad graphs

- Graphs should show something
- Too many slices ruin the pie chart, and be careful with color!
- The axis can distort the results

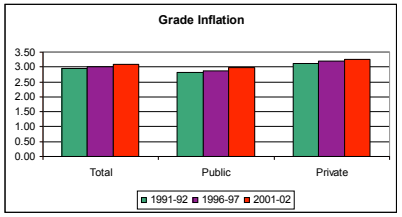


Is grade inflation a problem at universities?

- This is what was reported in an article at UD using national data.



- This is what I got after adjusting the axis to zero.



Some final thoughts

- Document!
- Use common sense!
- Graphs should help tell a story about your data - if there's no story the graph is wasting space and paper.
- **Unless, no results is the story!!!**