



Chapter 7 – Scatterplots, Association, and Correlation

Today's Data: Detroit Homicides

- The data are on the homicide rate in Detroit for the years 1961-1973.
  - <http://www.canynons.edu/faculty/morrowa/140/datasets/>

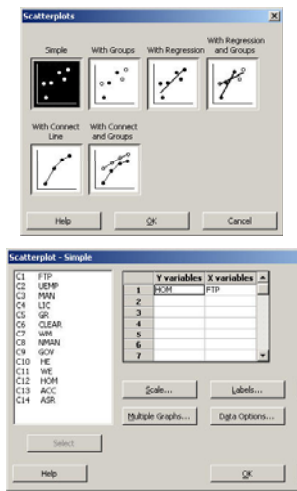
FTP - Full-time police per 100,000 population  
 UEMP - % unemployed in the population  
 MAN - number of manufacturing workers in thousands  
 LIC - Number of handgun licenses per 100,000 population  
 GR - Number of handgun registrations per 100,000 population  
 CLEAR - % homicides cleared by arrests  
 WM - Number of white males in the population  
 NMAN - Number of non-manufacturing workers in thousands  
 GOV - Number of government workers in thousands  
 HE - Average hourly earnings  
 WE - Average weekly earnings  
 HOM - Number of homicides per 100,000 of population  
 ACC - Death rate in accidents per 100,000 population  
 ASR - Number of assaults per 100,000 population

- Is there an association between the number of homicides and any other variable?

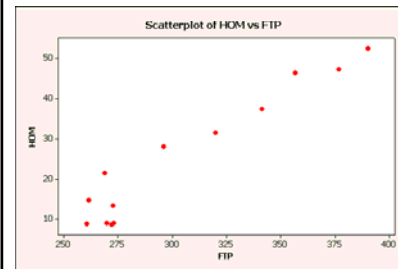
	FTP	UEMP	MAN	LIC	GR	CLEAR	WM	NMAN	GOV	HE	WE	HOM	ACC	ASR
1	260.35	11.0	455.5	178.15	215.98	93.4	558724	538.1	133.9	2.98	117.18	8.60	39.17	306.18
2	269.80	7.0	480.2	156.41	180.48	88.5	538584	547.6	137.6	3.09	134.02	8.90	40.27	315.16
3	272.04	5.2	506.1	198.02	209.57	94.4	519171	562.8	143.6	3.23	141.68	8.52	45.31	277.53
4	272.96	4.3	535.8	222.10	231.67	92.0	500457	591.0	150.3	3.33	147.98	8.89	49.51	234.07
5	272.51	3.5	576.0	301.92	297.65	91.0	482418	626.1	164.3	3.46	159.85	13.07	55.05	230.84
6	261.34	3.2	601.7	391.22	367.62	87.4	465029	659.8	179.5	3.60	157.19	14.57	53.90	217.99

Scatterplots

- A **scatterplot** provides a visualization of the relationship between two quantitative variables.
- Minitab > Graph > Scatterplot > Simple
  - Select a variable for the Y axis and a variable for the X axis
- Create a scatterplot for number of homicides per 100,000 of population and full-time police per 100,000 population

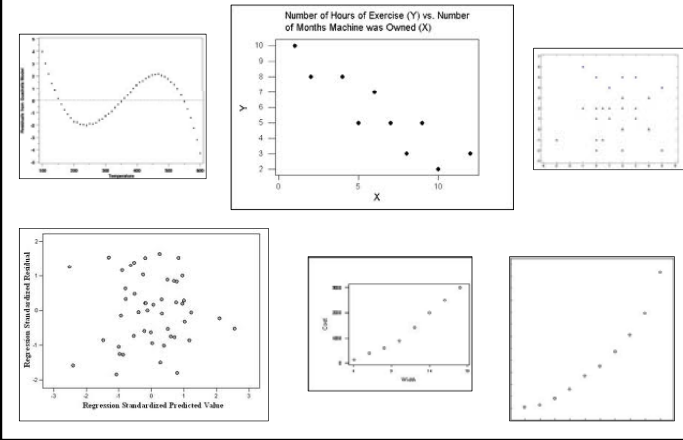


Just Checking...

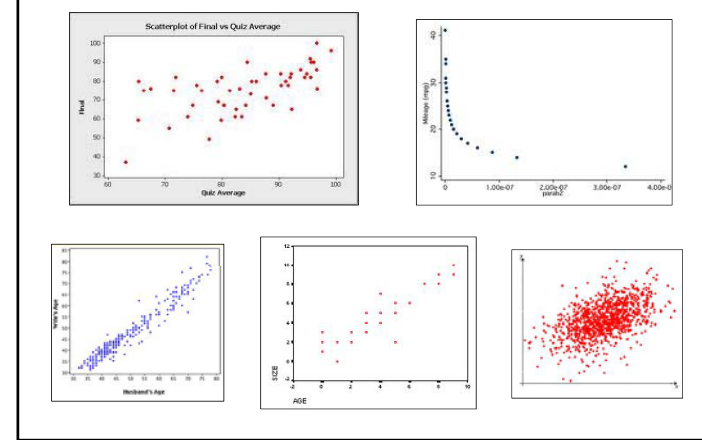


- Things to Note
  - Form
    - Linear (positive or negative)?
    - Other pattern?
    - No pattern?
  - Strength
    - How much scatter (from vague cloud to tight adherence to pattern)?
  - Unusual Features
    - Points far away?
    - Subgroups?
    - Nonconstant variance?

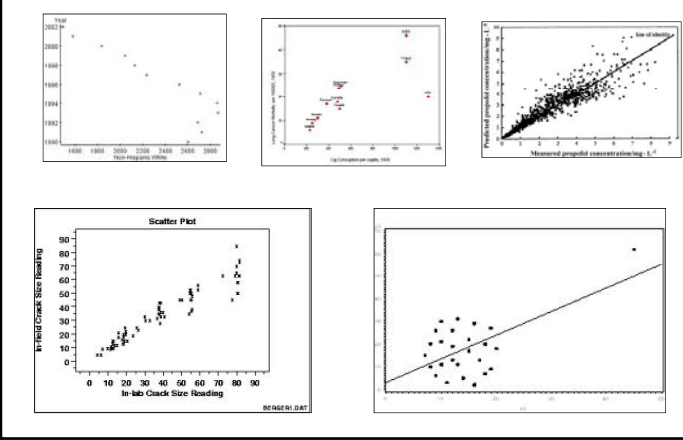
### Linear vs. Nonlinear vs. No Relationship



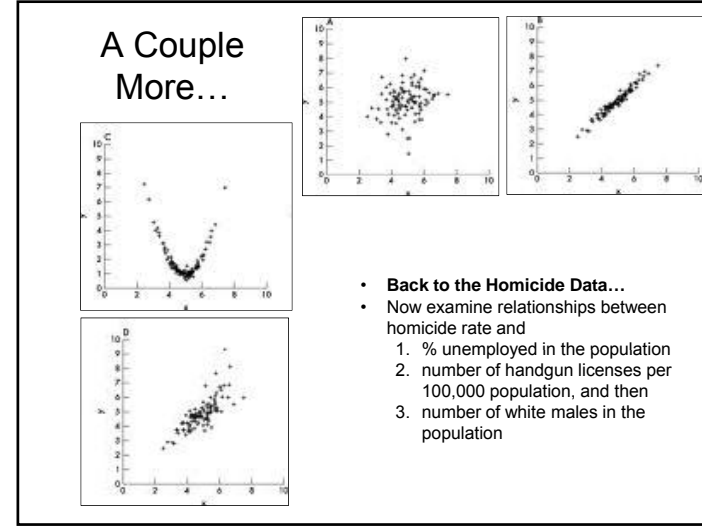
### Weak vs. Moderate vs. Strong



### Nonconstant Variance and Outliers

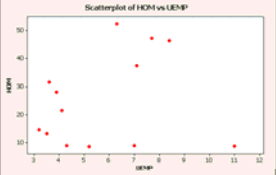
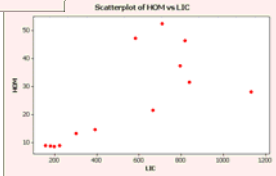
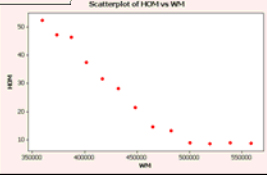


### A Couple More...



- **Back to the Homicide Data...**
- Now examine relationships between homicide rate and
  1. % unemployed in the population
  2. number of handgun licenses per 100,000 population, and then
  3. number of white males in the population

### Just Checking...

- Form
- Strength
- Unusual Features

The variable of interest is called the **response** and the other variables are called the **explanatory variables** or **predictors**.

## The Correlation Coefficient

- The **correlation coefficient**,  $r$ , measures the strength of a linear relationship between two quantitative variables.
- Conditions to Check Before Use
  - Quantitative Variables Condition
    - Both variables must be quantitative
  - Straight Enough Condition
    - Check the scatterplot – must look roughly linear
  - Outlier Condition
    - Outliers have potential to really change results
- If there's an outlier, report correlation with and without results


## So... How Do You Find $r$ ?

- To compute  $r$  by hand, we turn to z scores
 
$$r = \frac{\sum z_x z_y}{n-1}$$
  - $r$  doesn't change as we rescale/shift the  $x$  and  $y$  data
- Minitab > Stat > Basic Statistics > Correlation
  - Enter all variables
  - Uncheck box for p-values


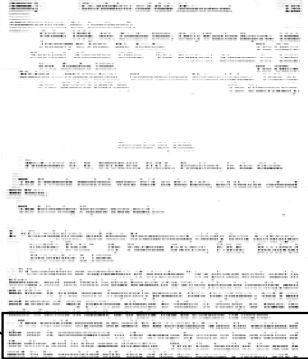
**Correlations: FTP, WM, HOM**

	FTP	WM
WM	-0.884	
HOM	0.964	-0.953

Cell Contents: Pearson correlation



## Galton: Origin of Correlation

- 1888 'Co-relations and their measurement, chiefly from anthropometric data.'
  - Proceedings of the Royal Society* 45 (December 13): 135-45
  - "Two variable organs are said to be correlated when the variation of the one is accompanied on the average by more or less variation of the other, and in the same direction."
- From <http://galton.org/>
  - "Galton was not himself a mathematician, though he was competent enough, but really an intensely practical man. He left the refinement of his techniques to dedicated and supremely talented mathematicians like [Karl] Pearson, but provided the crucial formation of concepts; he was a practical innovator, not a technician

### Correlation Properties

- $r$  measures strength of *linear* association ONLY
  - May be other types of association (quadratic, log, etc)
- $-1 \leq r \leq 1$ 
  - $r > 0$ : Positive linear association
  - $r < 0$ : Negative linear association
- Correlation treats  $x$  and  $y$  symmetrically
  - $\text{corr}(x,y) = \text{corr}(y,x)$
- $r$  has no units
  - Always report as a decimal
- $r$  is unchanged by shifts/rescaling of data
  - Q: What if we change HOM to total number of homicides?
- Correlation is very sensitive to outliers
  - Outliers can change correlation by A LOT
  - Always report correlation with and without outliers
- ALWAYS show a scatterplot when reporting  $r$

### Anscombe's Quartet

- Each picture has the same correlation:  $r = 0.816$ . Where is  $r$  appropriate?

### Correlation vs. Causation

- Correlation does NOT mean causation
- **Ex:** Ice cream consumption and drowning have a high correlation
- A **lurking variable** is a hidden variable that stands behind a relationship and determines it by simultaneously affecting the other two variables.
  - ie: it is a variable that affects the variables being studied but is not included in the study
- Q: There is a high positive correlation between HOM and FTP. Does that mean increasing the full time police will increase the number of homicides?

### Class Work

- To get credit, it is your responsibility to get checked off.
  1. Chapter 7 Handout
    - Rules for checking answers: No Pens in the Front!!!

### Homework

- Textbook/Routine Homework
  - Due Next Week (25% chance of collection)
  - 1. Read Chapter 7
  - 2. Pg 186-193: #3, 5, 9, 11, 15, 17, 23, 25, 27, 35
- Project/Exploration Homework
  - Project #1