# *Introduction to Data Analysis and Graphs*

"If you torture numbers enough, they will confess to anything." John Kennedy

## *Overview*

Information is increasingly packaged as summary numbers and graphs, which means you need to "push" numbers to make good decisions, whether it is buying a car or a home or deciding on a major. Ignorance is no excuse, so you need to make an investment in some rudimentary data analysis skills. It is true "there are lies, damn lies, and statistics," and you need to protect yourself from those liars. We do not go into the details of how to create numbers here because I do it in ECN201, but you should know someone "creates" those numbers / graphs / statistics you see and for those who are interested in a more thorough discussion of these, you might want to check out my ECN306 class that I teach every Spring.

Here we'll start with a quick look at some graphs because in this course you ***must*** read and interpret tables and graphs and you ***must*** be able to find the 'story' embedded in them. You will also be introduced to an important algebra technique that allows us to make sense of decisions such as the decision to go to college that involve future benefits and / or costs.

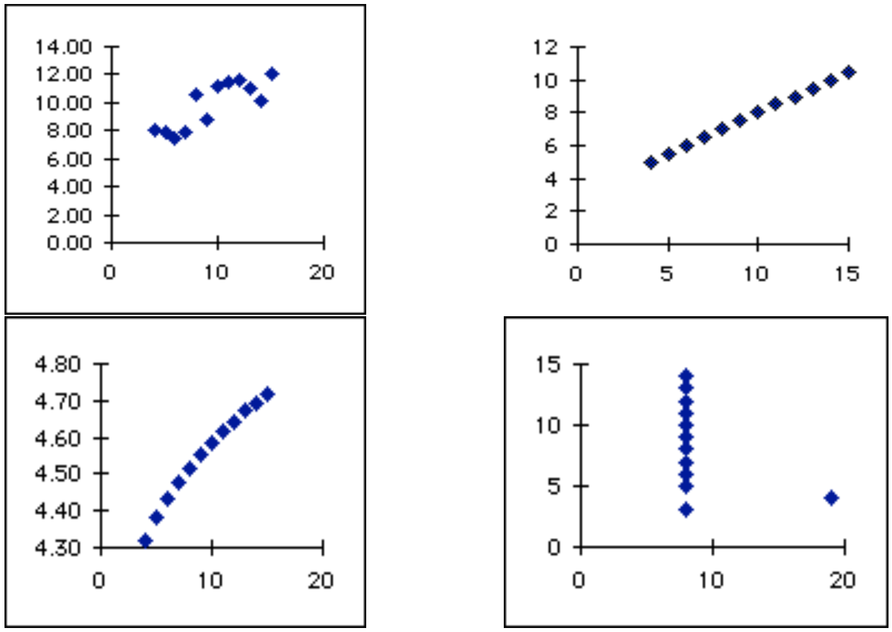
## *Graphs*

You have certainly heard the expression; a picture is worth a thousand words, and while this may be an overstatement, it is often useful to describe relationships in visual form. When you see a graph, you should think that behind each graph is a table of numbers and the creator of the graph was trying to make it easier for the reader to see the story in the numbers. For example, consider the following table of four relationships between the variables X & Y. Give yourself 30 seconds to review the table and write down a brief description of each relationship.

**The Table**

| | #1 | | #2 | | #3 | | #4 |
|---|---|---|---|---|---|---|---|
| X | Y | X | Y | X | Y | X | Y |
| 4 | 8.05 | 4 | 5 | 4 | 4.32 | 8 | 3 |
| 5 | 7.93 | 5 | 5.5 | 5 | 4.38 | 19 | 4 |
| 6 | 7.38 | 6 | 6 | 6 | 4.43 | 8 | 5 |
| 7 | 7.86 | 7 | 6.5 | 7 | 4.48 | 8 | 6 |
| 8 | 10.63 | 8 | 7 | 8 | 4.52 | 8 | 7 |
| 9 | 8.74 | 9 | 7.5 | 9 | 4.55 | 8 | 8 |
| 10 | 11.15 | 10 | 8 | 10 | 4.58 | 8 | 9 |
| 11 | 11.52 | 11 | 8.5 | 11 | 4.62 | 8 | 10 |
| 12 | 11.64 | 12 | 9 | 12 | 4.64 | 8 | 11 |
| 13 | 11.09 | 13 | 9.5 | 13 | 4.67 | 8 | 12 |
| 14 | 10.17 | 14 | 10 | 14 | 4.70 | 8 | 13 |

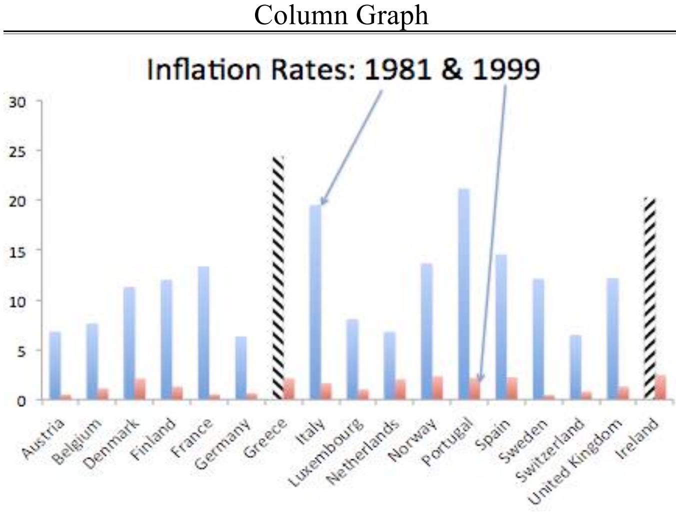
Once you have done that, look at the four graphs and write down a brief description of each relationship – the "story" you extracted from the graph. The fact is the graphs and the table contain the same information, and while both are correct, in this case I would use the graphs to represent the relationships.



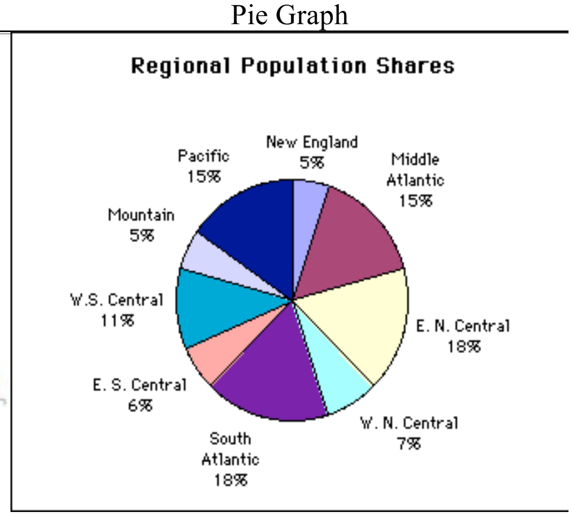
The graph types we focus on in this course are the Line, Scatter, and Time-series graphs, although as college students you should be able to interpret pie and column / bar graphs. Every time you see a graph you should focus on extracting the story embedded in that graph. If a picture is worth 1,000 words, then you should be able to reverse engineer any pictures back into the corresponding words. It is the ability to translate tables and graphs into words that will prove to be an invaluable skill, so get some practice now because reading graphs is like riding a bike – it is difficult in the beginning and you might fall a few times – but with practice you can master it and then it seems very easy. All graphs are pictures of relationships – so start with figuring out what are the variables in the relationship. Also make sure you recognize the difference between describing a graph and explaining it. Describing it usually involves phrases like “it was rising” or “they are inversely related.” You are simply describing the pattern you see and once you have a good idea what it is you are measuring, it's time to focus on describing what you see. Start with anything striking about the graph. Explaining is more difficult because you are then trying to identify what might explain the pattern you are seeing, and this usually means you need to have some outside information. In this course you will be asked to do both.

Below you will find four graphs that you should spend a minute and work on the story behind each picture. Begin with a good description of what you “see” and then take a shot at explaining them.

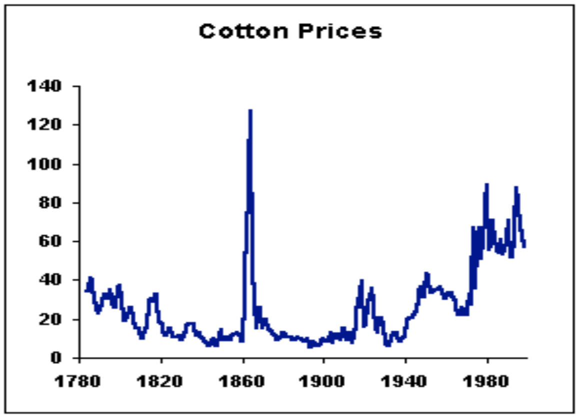
Column Graph

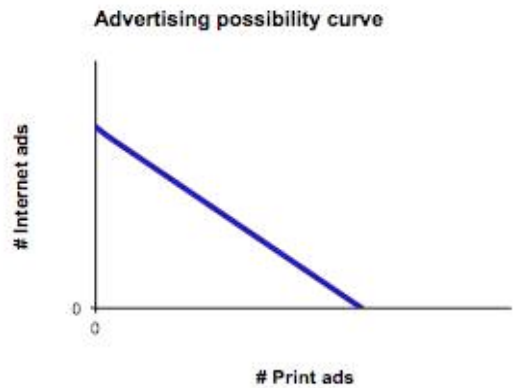



Pie Graph



| Time-series Graph | Line Graph |
| --- | --- |




In the column graph this is a story about inflation rates in wealthy countries, mostly in Europe, and how they changed between 1981 and 1999. Two things are very obvious. First, in 1981 inflation rates varied greatly across the countries with some inflation rates above 20% a year and many with inflation rates above 10%. Second, inflation rates were sharply lower in 1999 across all of the countries and the variation among the countries was also smaller. What is the explanation here? It turns out that many of these countries were working toward adopting the euro as their currency and one of the requirements was to get the national inflation rate below a specified target. The decline we see here is also true in other countries as the world experienced a downward trend in inflation rates since the late 1970s.

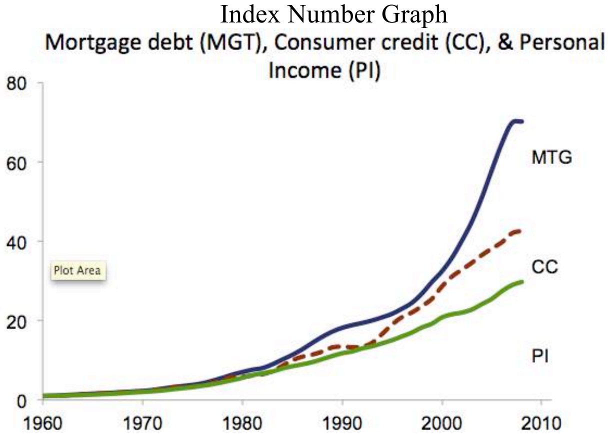
In the time-series graph we are looking at how cotton prices have changed over time. The most obvious feature of the graph is the spike in the early1860s, but there are some other interesting price movements. Look for long-term trends or shorter cycles. In this case there tends to be a downward trend into the late 1850s and an upward trend that begins in the 1930s. There also are some notable cyclical swings - a spike between 1810 and 1820, another one around 1920, and one in the 1940s. To move beyond description, you need to know some history. For example, it looks to me like these blips are related to wars - the War of 1812, US Civil War (early 1860s), WWI (late 1910s), and WWII (early1940s). You also see a sharp increase in the 1970s, a decade of rapidly rising inflation rates.

The final graph is a line graph. These are very important in economics AND they are very difficult to interpret. In fact you will find very few of these in magazines because so few people can master them, but you are college students and you should be in the group that can interpret these graphs. This is a graph describing the relationship between the number of Internet ads and the number of print ads a company can afford with its ad budget. A key feature of the graph is the slope of the line: the slope is negative and this means the two variables are negatively related. If the company spends more on its Internet ads then with the same budget it will be able to buy fewer print ads.

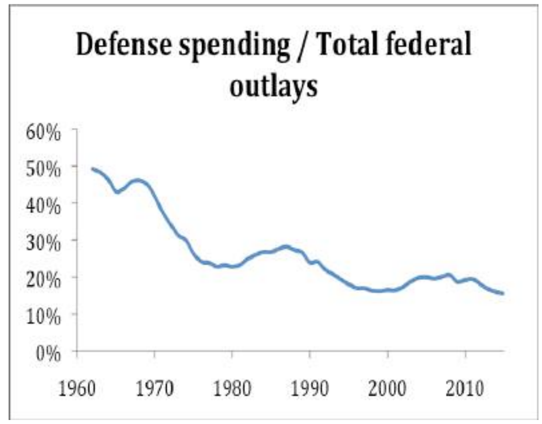
In each of the above graphs the variables being measured were obvious – time, country, population – but sometimes the variables are a bit less obvious. Below you will find two such examples. In the first graph there are *index variables* and in the second there are *ratio variables*. When you see these variables keep in mind that you lose information on the actual numbers. For example, in the two graphs below you cannot provide any information on the actual value of mortgage debt in 2000 or the level of defense spending in 2010. Index and ratio variables are used often to help highlight the relationship between two variables over time.

In the Index Number Graph there are data for two types of debt – mortgage and consumer credit – and personal income. You know you are looking at an index number if the various lines all start at 1 at some time. The advantage of the index approach is it provides a better visual representation of comparative growth, while its disadvantage is the lay reader without some guidance does not readily understand the index number. In this graph we can see that over time all of these variables have increased, although there seems to be a difference between growth before and after 1980. Before 1980 all variables tended to move together, but after 1980 debt expanded more rapidly than income. You also see that the growth in mortgage debt "exploded" after 2000, and in the course's final unit we will look into an explanation for this pattern.

Index Number Graph



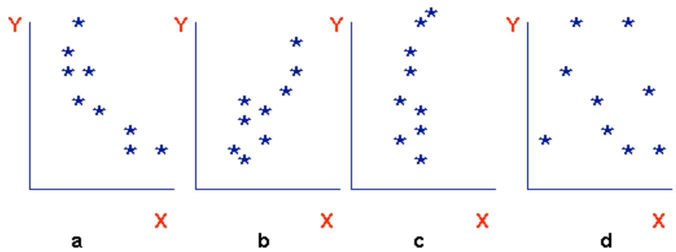

Ratio Variable Graph




In the case of the Ratio Variable Graph you ***CANNOT*** say anything about the level of total spending or defense spending. It is impossible to see if defense spending increased or decreased or when defense spending was highest or lowest. All you can see is how the allocation of spending changed over time. Looking at the defense-spending graph, the most notable feature is the general downward trend in the share of federal outlays allocated to defense. There were, however, also three periods where the decline was reversed - the late 1960s, the early 1980s, and in the early 2000s. In this sixty-year period defense spending's share of federal outlays dropped from about 50% to nearly 20%. To move from description to explanation, you need some history. The positive "blips" correspond to the Vietnam War (late 1960s), the Reagan defense buildup (early 1980s), and the Bush "War on Terrorism (early 2000s)," while the overall decline corresponds to the growth in entitlement programs such as Social Security and Medicare that are taking up an increasing share of the budget. Also behind the general decline is the demobilization following Vietnam (1970s), and the collapse of the Soviet Union that ended the Cold War and created the "Peace Dividend" (1990s).

The scatter diagram has caused students the most problems over the years, although it need not be this way since there is nothing difficult about interpreting a scatter graph designed to provide a visual image of the nature of a relationship between any two phenomena. For example, assume you have decided to undertake a study to determine the relationship between two phenomena, X and Y, and as a first step you collected nine observations. What you want to know is whether or not these data support the hypothesis there is a relationship between y and x.

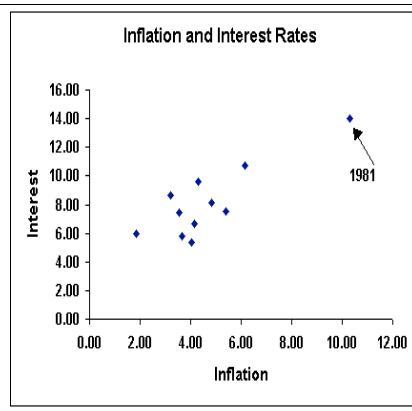
Below are four possible 'patterns' that could emerge from the analysis. In each diagram the points correspond to the individual years, and you are looking for a pattern in the data. In the first diagram the points tend to be loosely scattered around a *negatively* sloped line, which provides evidence that y and x are *negatively* related. This might be what I would find if I surveyed the class to find the relationship between number of hours spent on social media and semester GPA. In the second diagram, meanwhile, the points seem to be more packed around a *positively* sloped line suggesting a *positive* relationship between x and y – what I might find in a class survey of average weekly study hours and semester GPA. In the third graph the scatter of points resembles a vertical line suggesting that Y can take on any values while X does not change much, and in the fourth graph there is little evidence of any relationship.




Before leaving scatter diagrams behind, let us turn to the specific problem. What is the relationship between inflation and interest rates? Economic theory leads you to believe interest rates (r) and inflation rates (i) are positively related - an increase in inflation rates pushes up interest rates. To test this theory the data on interest rates and inflation in the table below were collected. Do these data support the hypothesis there is a relationship between interest rates and inflation? To answer this we'll look at the scatter diagram associated with these data.

There does appear to be a relationship between the two variables since the scatter of points tends to rise as we move to the right. The data support the theory that the two variables are positively related - as the inflation rate increases, interest rates tends to increase also.

| *Year* | *Interest Rate* | *Inflation* |
|---|---|---|
| 1981 | 14 | 10.3 |
| 1982 | 10.7 | 6.2 |
| 1983 | 8.6 | 3.2 |
| 1984 | 9.6 | 4.3 |
| 1985 | 7.5 | 3.6 |
| 1986 | 6 | 1.9 |
| 1987 | 5.8 | 3.6 |
| 1988 | 6.7 | 5.8 |
| 1989 | 8.1 | 4.8 |
| 1990 | 7.5 | 5.4 |
| 1991 | 6.1 | 3.9 |




And do not forget to master line graphs since you will be responsible for them this semester.

Now we'll look at an important application of algebra for macroeconomic - the conversion of actual, or nominal values, into real, or inflation-adjusted, variables.

## *Inflation-adjusted Data*

When using any variables measured in dollars - income, earnings, sales, profit, and GNP - you ***MUST*** be careful when interpreting changes in these variables over time. For example, if you received a wage increase of 20%, how would you feel about it? Would this be good news? Would you now be in a better financial position?

It turns out that you do not have enough information to answer that question. If your wages rose 20% while prices rose 40%, the buying power of those wages would fall, while if prices rose 1o%, the buying power of those wages would rise. We need a technique that allows us differentiate these two situations. We need a way to correct for the distortion caused by rising prices; if you are interested in changes in the buying power of workers' wages, you need to account for and changes in the prices they pay.

Fortunately, we have one. To capture this effect, economists have developed such a technique. When looking at any dollar-denominated variable over time, economists construct a new variable known as the ***real, constant dollar,*** or ***inflation-adjusted*** variable. Regardless of what you call it, the concept is straight forward enough.

First, we need a measure of prices to use, and in most instances the Consumer Price Index (CPI), which we examine in some detail in a later unit, is used.

Second, we create a new variable - the 'inflation adjusted', or Real wages variable using the formula below. *Real* (inflation-adjusted) wages are simply *Nominal* (actual) wages divided by the price level.

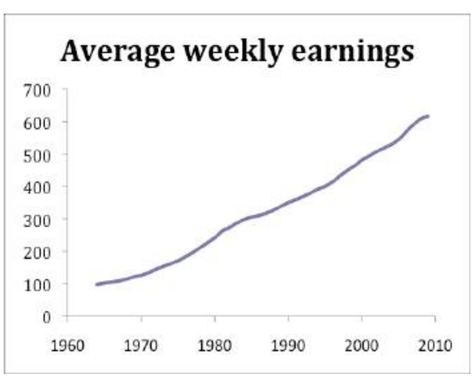
**R = N/PI *100**

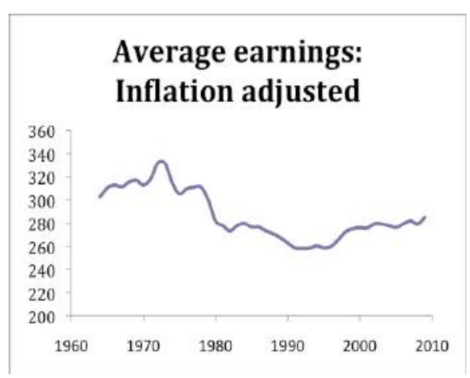
- *R* = real value (constant dollar)
- *N* = nominal value (current dollar)
- *PI* = price index

To see how the approach works we begin with some data on earnings and the price level (CPI). In the second column we have the average weekly earnings for a worker in the US and we see that in the fifty-year period earnings have risen by more than a multiple of 9 (480/53 = 9.05). In the third column is the price level and we see that prices have increased by a multiple of more than 7 (172/24 = 7.14). Based on this we should expect inflation-adjusted data on real earnings to reflect some growth since earnings rose faster than prices.

| | *Earnings* | *CPI* | *Real Earnings* |
|---|---|---|---|
| 1950 | $53.13 | 24.1 | $220.46 |
| 1960 | $80.67 | 29.6 | $272.53 |
| 1970 | $125.80 | 38.8 | $324.23 |
| 1980 | $240.77 | 82.4 | $292.20 |
| 1990 | $349.29 | 130.7 | $267.25 |
| 2000 | $480.41 | 172.2 | $278.98 |

The importance of the adjustment is evident in the graphs below entitled Average Weekly Earnings. In the left-side graph, the untrained eye sees continual improvement in average weekly earnings. Furthermore, given the fact that earnings increased at an average yearly rate of 4% in both the 1960s and 1980s, and 7% in the 1970s, you might be led to believe the 1970s was a period of more rapid growth. But you did not adjust for inflation, and when you do the picture is very different. Real earnings of American workers peaked in 1973 and by 1990 they had fallen to 1960 levels. Yes wages increased 7% per year in the 1970s, but prices increased nearly 8% giving us an average yearly 'decline' of 1% in wages. This diagram also helps explain the proliferation of reports circulating in the 1990s indicating the plight of Generation X, those who moved into the labor force in the early 1990's. The forecasts that this generation might be the first to not achieve a standard of living higher than that of their parents is simply the result of an extrapolation of trends in the 1980s. Fortunately the trend reversed itself for most of the 1990s, although the decline returned after 2000.



A similar problem exists when we examine interest rates. Consider the position of a moneylender who must determine the appropriate interest rate to charge. One of the considerations will be the rate of inflation, the rate of increase in the price level (CPI). If the inflation rate is 6%, a lender must receive 6% interest just to maintain the money's buying power. If the cost of living increases 6% this year, then what you buy today for $100 will cost $106 next year, so a lender charging 6% on $100 will get $106 at the end of the year - just enough to break even. If on the other hand, the lender wanted a 2% return on money, then the interest rate would need to be 8%, with 6% simply accounting for inflation.

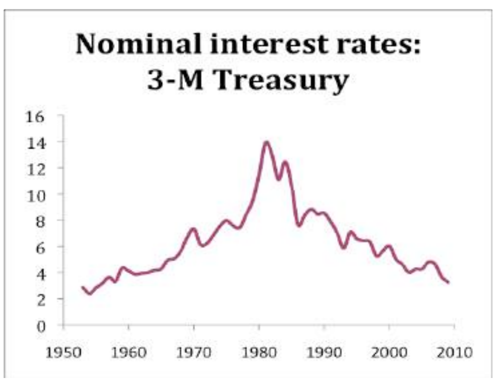
The realization that inflation rates are a common denominator in interest rates has prompted economists to develop a concept called 'real interest rates'. The unobserved 'real' rate, which is what 'really' matters to decision makers, is defined as the actual rates minus the expected inflation rate. The relationship between real and nominal rates is given by the following equation.

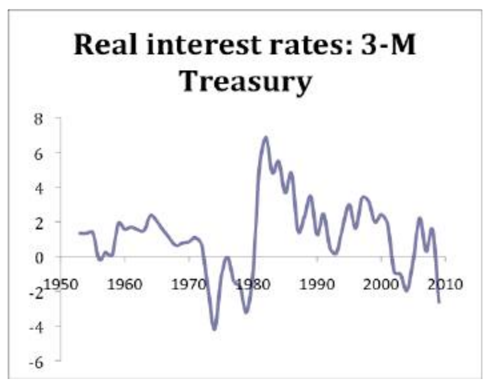
$$r_r = r_n - i$$

where:

$r_n$ = actual interest rate (what you see in the news)
$r_r$ = real interest rate
$i$ = inflation rate

As with wage earnings, there is a significant difference between the movement in real and nominal interest rates that is evident in the graphs below. In the 1980s nominal short term rates on government securities fell sharply from 11.5% in 1980 to 6% in 1986 suggesting lower borrowing costs, but that was not the case because inflation rates dropped even faster. As a result real interest rates actually rose in the early 1980s.




The bottom line is that inflation matters and that financial data, numbers expressed in dollar quantities, need to be adjusted for changes in the price level. Now we will move on to a discussion of benchmarking, adjustments for scale differences.