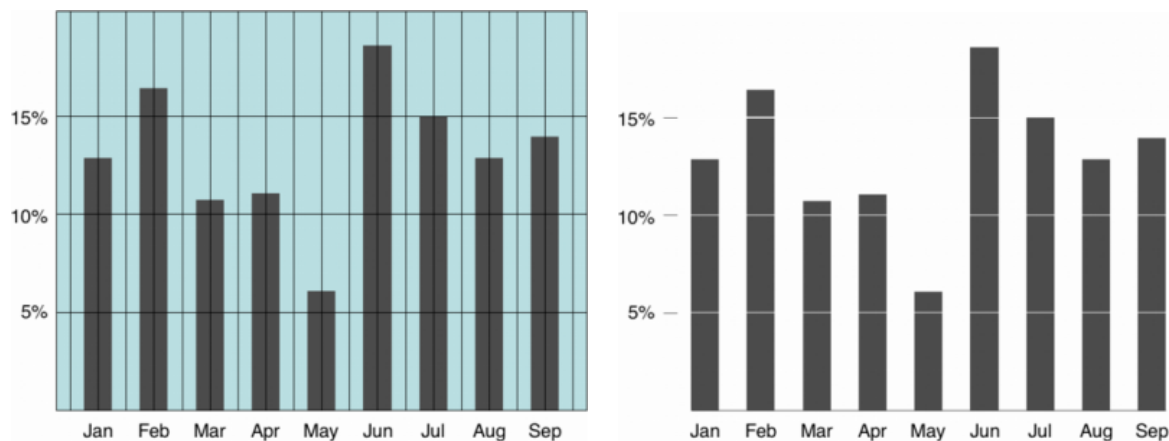# Tufte and Wainer

We're also going to talk today about Tufte's and Wainer's ideas on graphical presentation.
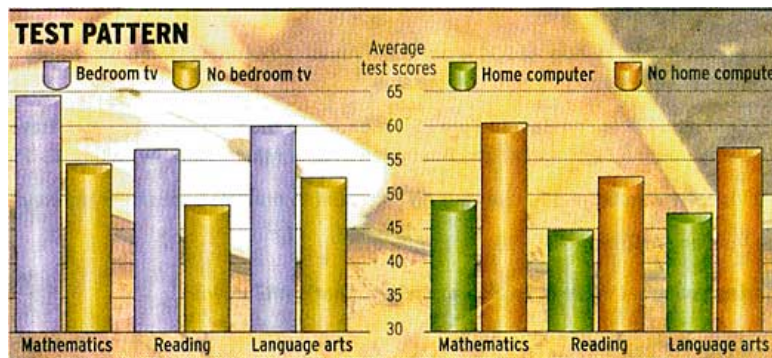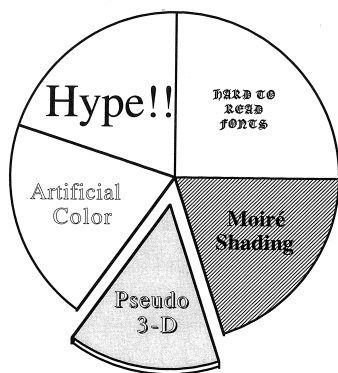
"Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space." — Edward Tufte

We'll also talk about Tufte's ideas on "data-ink." Compare these two plots.

And his campaign against "chartjunk." Have you seen plots that use chartjunk? How could this plot be improved?
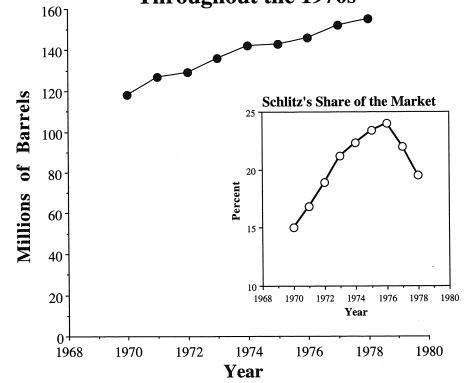
We also need to be careful for plots where the dimensions do not match the data. What is misleading about the left plot?

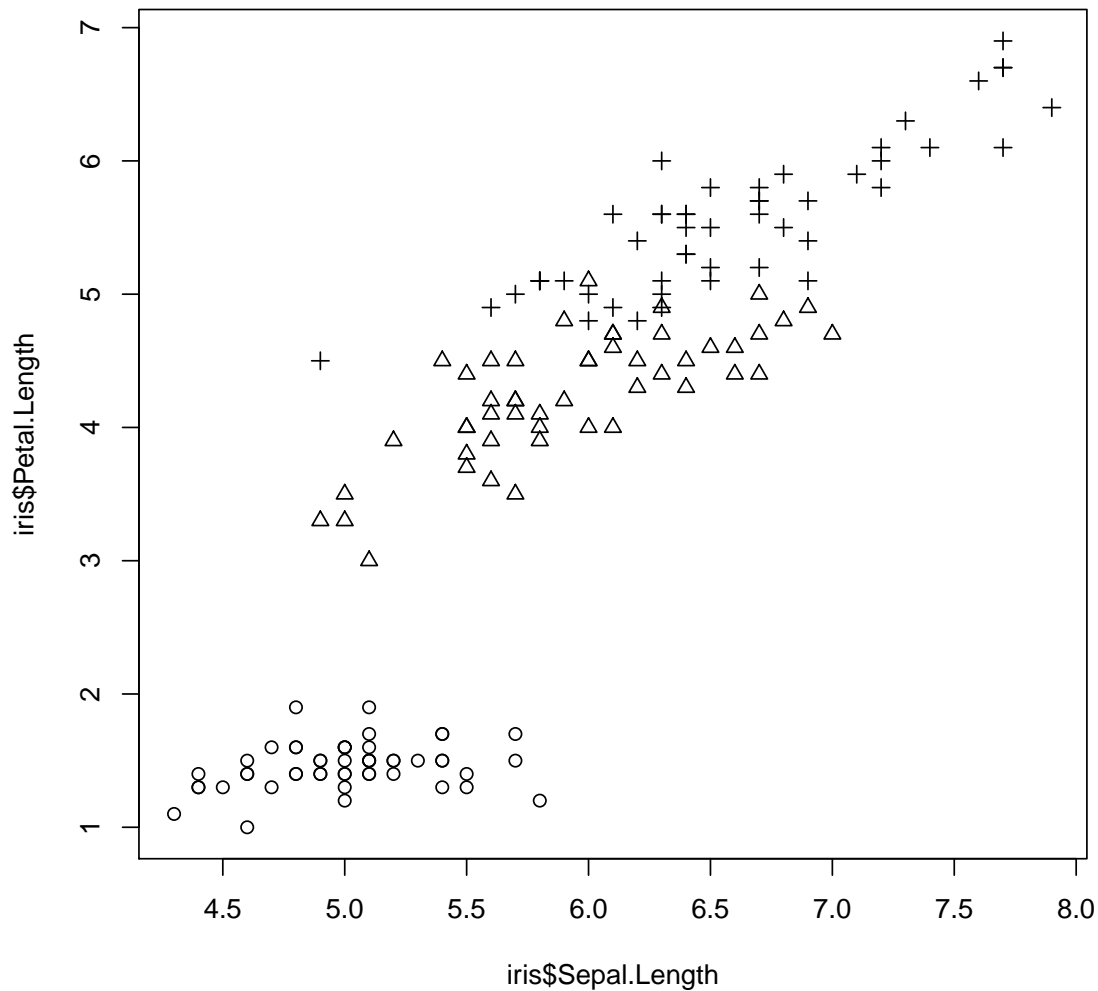### U.S. Beer Sales and Schlitz's Share

Barrels show total U.S. Sales

And to make sure our plots actually visually compare what we want to show. These three plots have the same information. What is the story, and which makes it clearest?
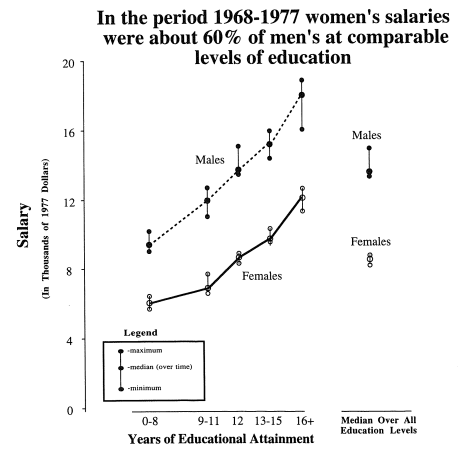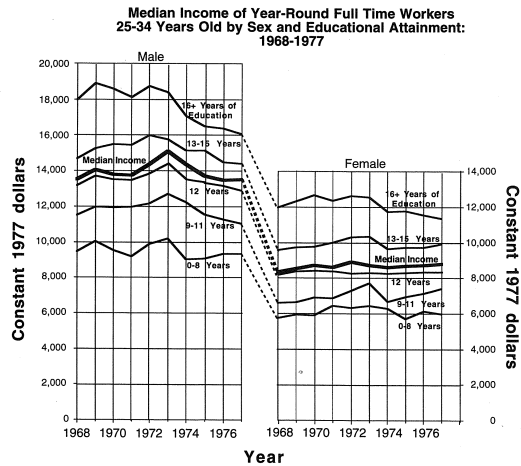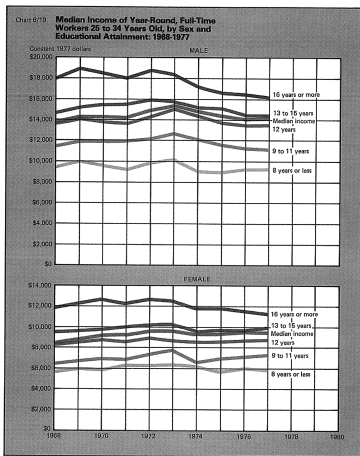
How should this plot be labeled and captioned? Fix it as best you can.

What else might you change?

BEWARE OF DYNAMITE

Tatsuki Koyama, PhD
Division of Cancer Biostatistics
Department of Biostatistics, Vanderbilt University School of Medicine

Introduction

One of my professional pet peeves is dynamite plots. Sometimes they are incorrectly referred to as bar plots. Dynamite plots do not have a formal name because they are not a part of conventional statistical graphics that should be used in reporting scientific results. But they are everywhere!

Why Are Dynamite Plots Bad?

The height of the bar represents the average, and the whisker the standard deviation (or standard error). They do resemble the dynamite you may see in cartoons, don't they? Why are they bad?
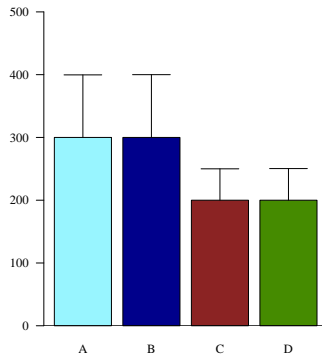
So little information! This plot presents 4 means and 4 standard deviations (or standard errors). This is a very inefficient use of space.

What do means mean? Averages do not usually convey much information. How spread are the data? Are there outliers? What are the sample sizes? None of these interesting and important questions is answered.

Whiskers get in the way. Whiskers add some information, maybe, but I don't know how to use them. They make the bars look taller, and the little information given by the bars is distorted.

Where are the data? There are data behind the bar, but the bar also covers space where no data exist. In addition, there must be some data above the bar in all that blank space.
Where are the data? You learn very little about the data and their *distribution*.

What might we conclude from this dynamite plot?
• Groups A & B are identical. • Groups C & D are identical. • Groups A & B are more spread than C & D.

What Should We Use Instead of Dynamite Plots?

We ought to show every single data point unless the sample size is too big. In that case, some summary measures may be used, but very rarely is a single average sufficient to represent the entire dataset. I like showing every single data point as a dot. Alternatively, a box plot showing the median, quartiles, and outliers may be used for a larger dataset.
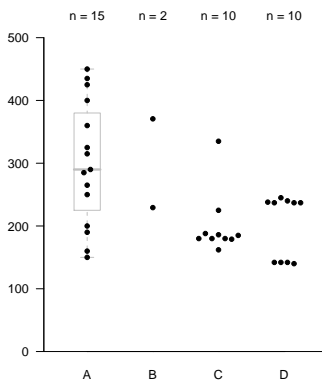
Here's the same data set shown in a way that reveals the individual points and their *distributions*.

• Are groups A & B identical?

• Are groups C & D identical?

• Are groups A & B more spread than C & D?

Furthremore, why are there only 2 data in group B? What is special about the one datum in group C that stands alone at the top? Are there two distinct groups within group D? What is going on?

Obviously, this is a made-up example. Mean-spirited, I might add. The fundamental problem with the dynamite plot is that it allows such a malicious example to be constructed. If all data points are shown, this is not possible. Intentionally or not, a dynamite plot hides more than it reveals!

Questions / Comments You Might Have

"I have used dynamite plots for the last 30 years, and I have a successful publishing record. Why try something new?"
Because *this* something new is the right thing to do. Show the data, don't hide them!

"I have never seen 'dot plots' in my favorite journal. My paper will be rejected if I use them."
I have never had problems publishing papers with this type of plot.

"But everyone is using dynamite plots..."
If everyone jumped off a bridge, ...

"But your plot is so busy."
It's not *too* busy, whatever *that* means. My plot may look busy to you, because it contains information. I'm all for simplicity, but quality of data presentation should not suffer to achieve simplicity.

"All right. I will give it a try. Oh wait... my favorite software only makes dynamite plots."
Maybe you should ditch the dynamite-producing software. You can create plots to show individual data and their *distribution* at http://data.vanderbilt.edu/~graywh/dotplot/.
Maybe you should work with a statistician. If you don't know where to find one, the biostatistics clinics is a good place to start. (http://biostat.mc.vanderbilt.edu/Clinics)

In Closing

Data are precious and usually expensive; treat them nicely. **Respect them as individuals!**

tatsuki.koyama@vanderbilt.edu

STAT8801, February 22, 2017

Name: ———————————————

We discussed some of the research about graphics by Cleveland and Tufte today.

What did you find particularly interesting or new to you?

————————————————————————————————

————————————————————————————————

————————————————————————————————

————————————————————————————————

————————————————————————————————

STAT8801, February 22, 2017

Name: ———————————————

We discussed some of the research about graphics by Cleveland and Tufte today.

What did you find particularly interesting or new to you?

————————————————————————————————

————————————————————————————————

————————————————————————————————

————————————————————————————————