Intro to Social Science Data Analysis

Lecture 7: Overview of Statistical Inference

Christopher Gandrud

October 8, 2012

Assignment 2

Due: Friday 19 October

Describe at least **3** variables in a data set.

You need to select a **range of descriptive statistical tools**. The tools should include both **numerical descriptive statistics** and **graphics**.

These tools should describe the variables':

- central tendency,
- variation,
- their relationships with the other variables.

The descriptions need to be discussed **in paragraph form**.

The description must be **reproducible**. So you should email me the link to a Dropbox folder with:

- the `.csv` data set,
- the `.Rmd` R markdown file,
- the final `.html` file.

Principles of Quantitative Graphics (1)

What are some of the goals that quantitative graphics should try to acheive?

Principles of Quantitative Graphics (2)

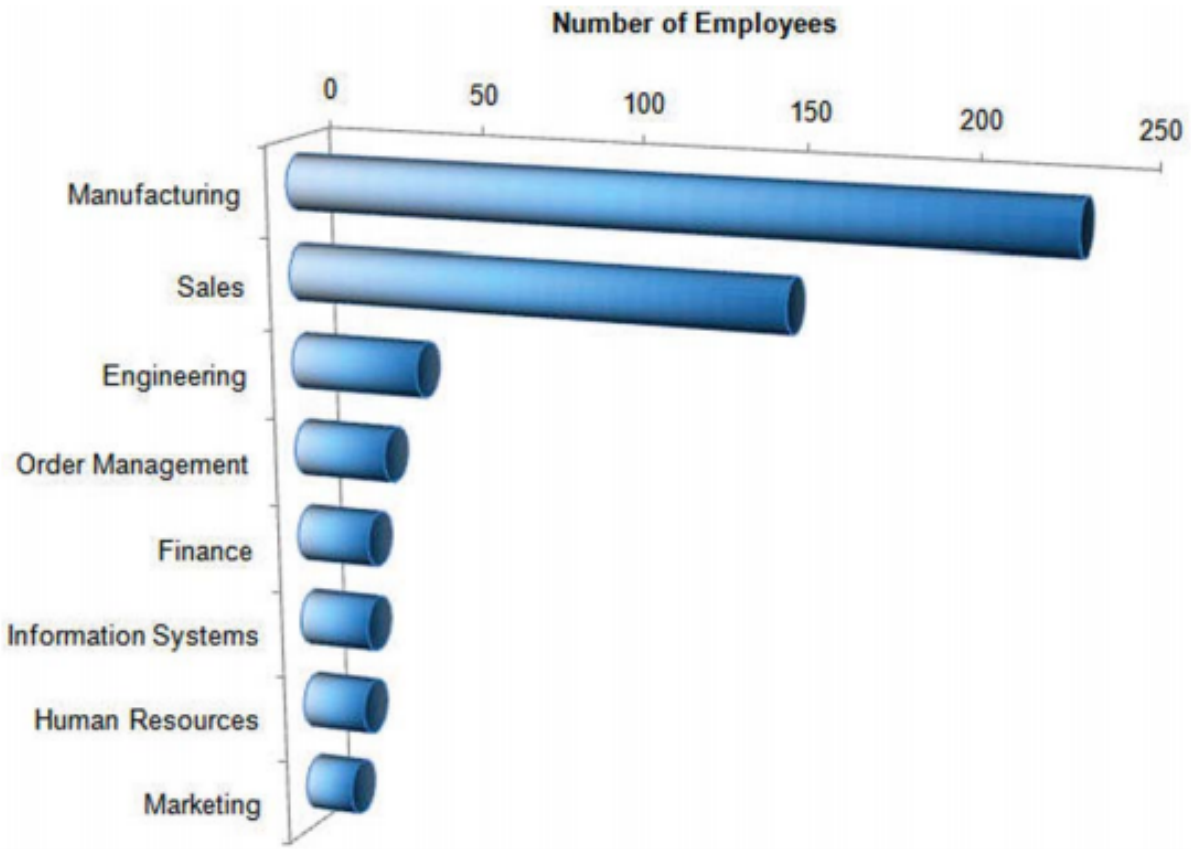
What are some things you should not do in quantitative graphics?

In particular, how does the **data-ink ratio** help us create effective graphics?

Remember:

$$\text{Data Ink Ratio} = \frac{\text{data} - \text{ink}}{\text{total ink}}$$

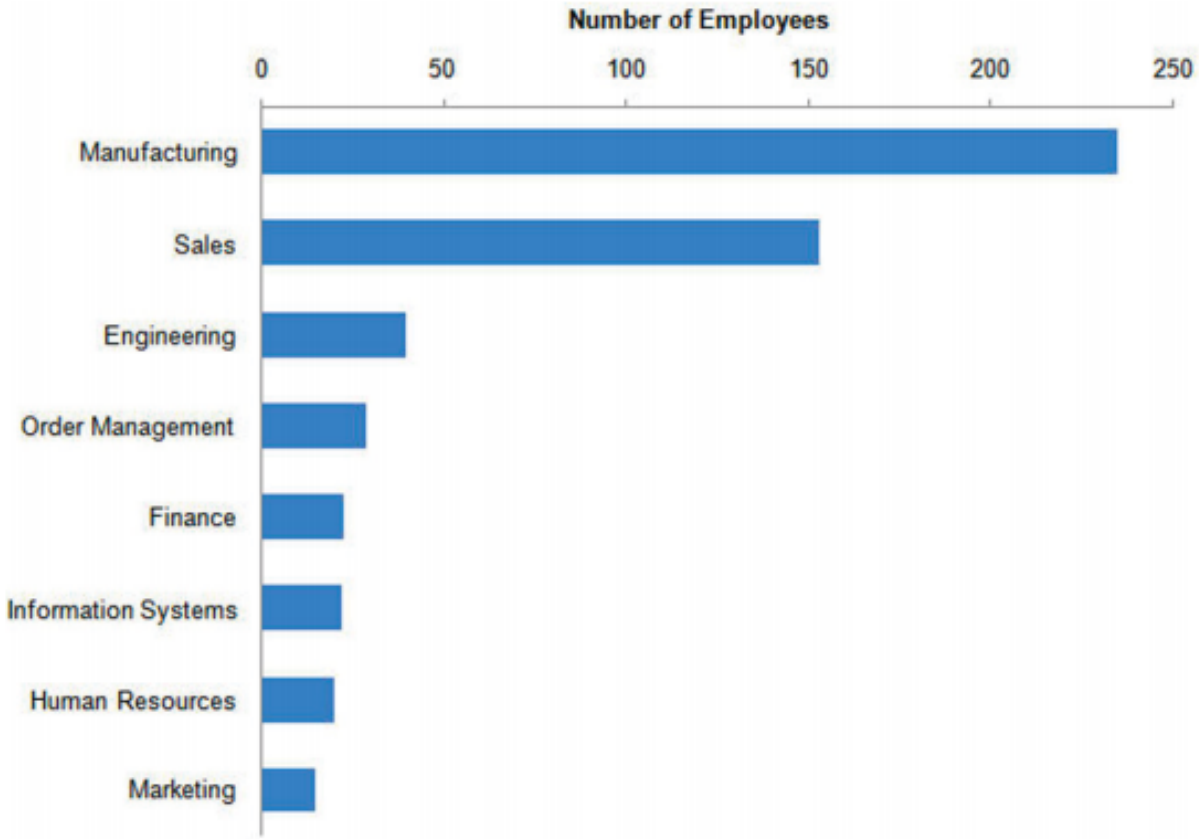
What is wrong with this graph?



Source:

`http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf`

Improved



Source:

`http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf`

Today's Data

Most of the examples for today's lecture will use data from the OpenIntro book.

You can download it like this:

```
 Load package RCurl library(RCurl)

## Loading required package:  bitops

 Download data Run10 <- getURL("https://raw.github.com/chri
 Convert Run10 to data.frame Run10 <- read.table(textConne
sep = "\t", header = TRUE)
```