Effective Plots and Tables

Jeffrey Michael Franc
MD, MSc, FCFP.EM, Dip Sport Med, EMDM

Medical Director, E/D Management
Alberta Health Services

Associate Clinical Professor of Emergency Medicine
University of Alberta

Visiting Professor in Disaster Medicine
Universita' Degli Studi del Piemonte Orientale

MedStatStudio

Objectives

- Know the advantages and disadvantages of various plots and how to choose the best plot to emphasize data

- Describe general rules for increasing the clarity of tables.

Plotting

Always plot the data in some meaningful way!!!!

Plotting

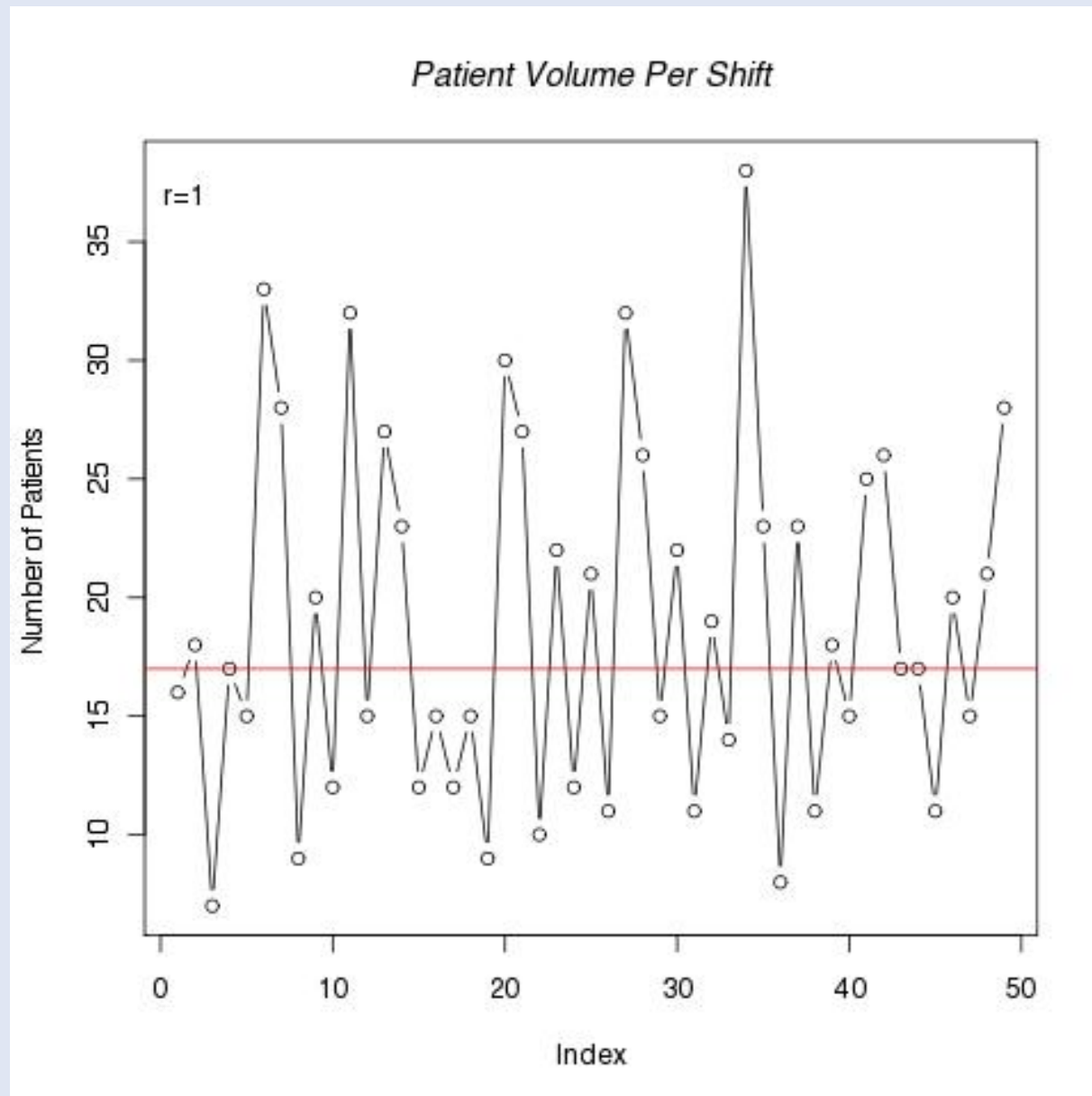Why should we plot the data?

Plotting

Why should we plot the data?

- Recognize unexpected trends
- Readers vary in sophistication
- P-values / Confidence intervals can be difficult to visualize
- Often many important features of the data can be expressed in a single plot
- Avoids monotony of continuous text

10 Favorite Plots

1. Run Chart

2. Scatter Plot

3. Density Histogram

4. Effect Plots

5. Barplot

6. Boxplot

7. Pareto Chart

8. Control Chart

9. Analysis of Means
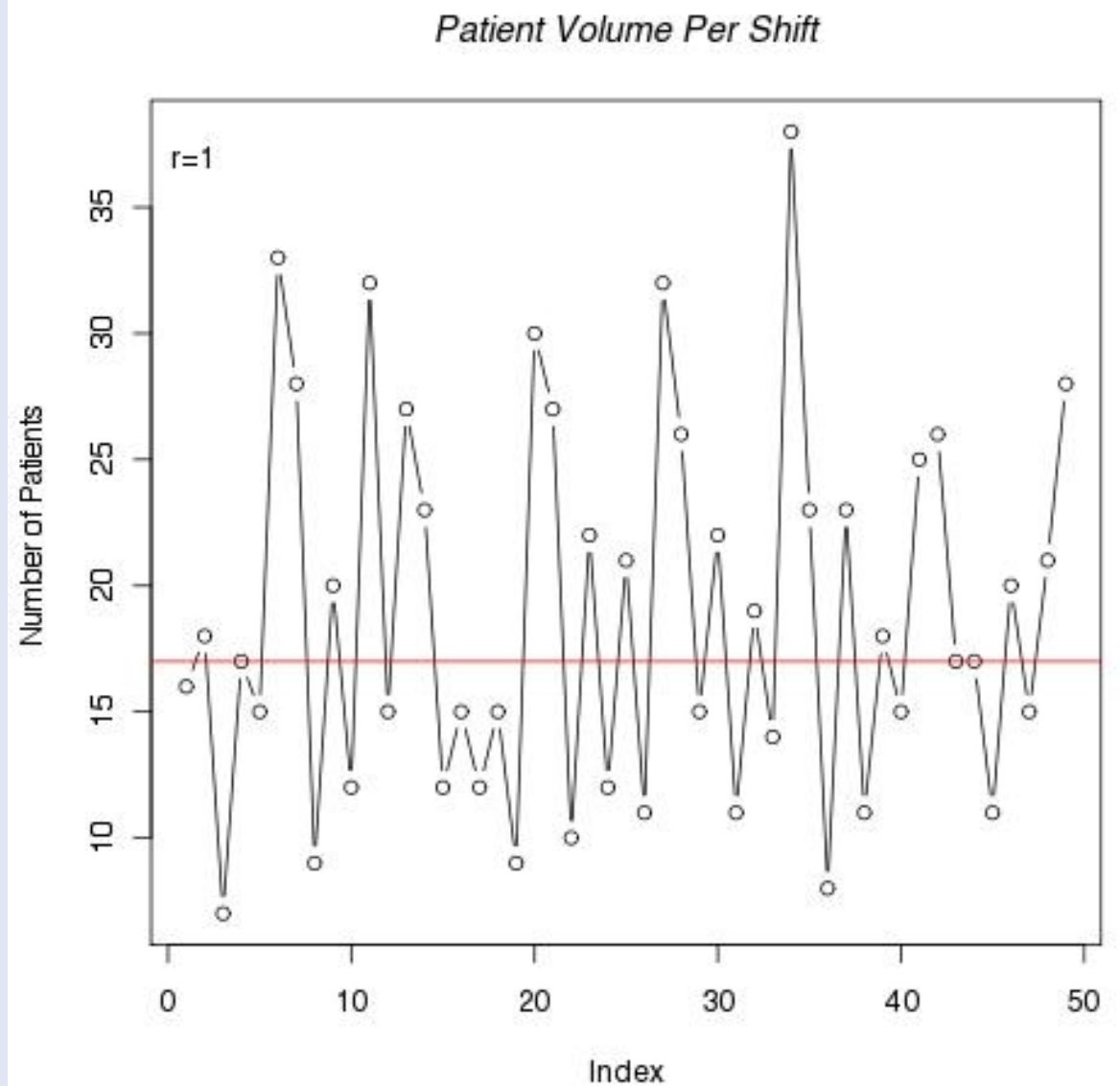
10. Residual Plots

11. ~~Pie Chart~~

1: Run Chart

Patient Volume Per Shift

Run Chart

Advantages

Disadvantages

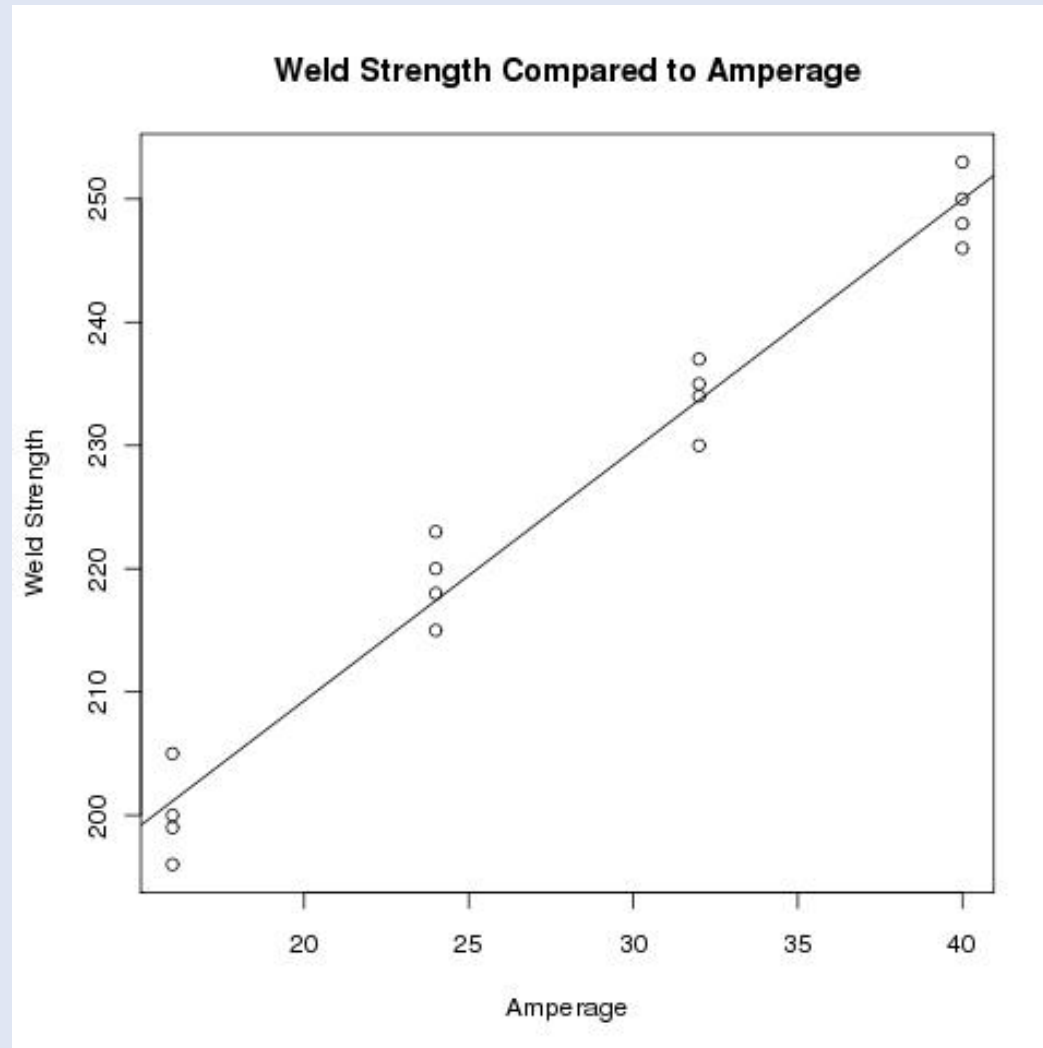Patient Volume Per Shift

Run Chart

Advantages

- Order of observations is preserved

- Good to assess for patterns
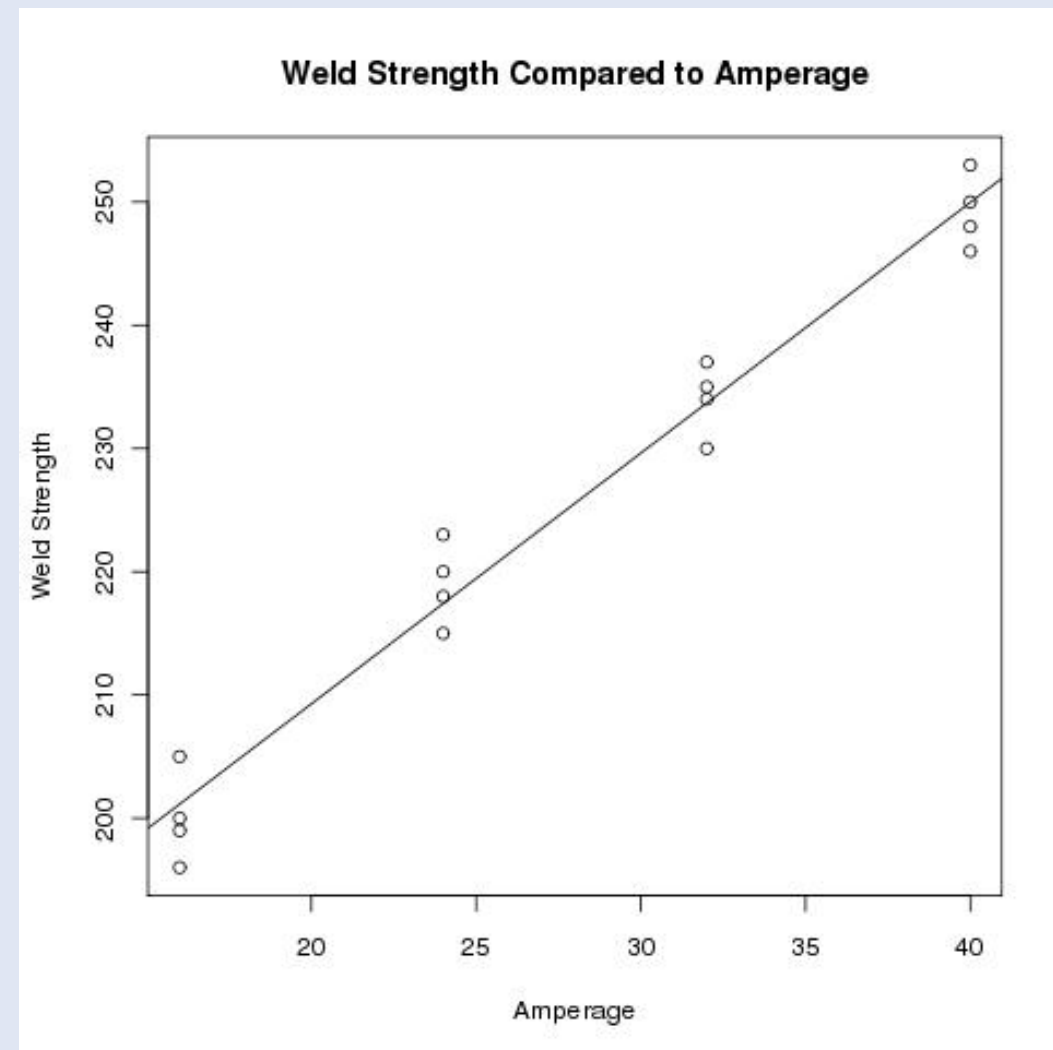
Disadvantages

- Sometimes order is irrelevant

2: Scatter Plot

Scatter Plot

Advantages

Disadvantages
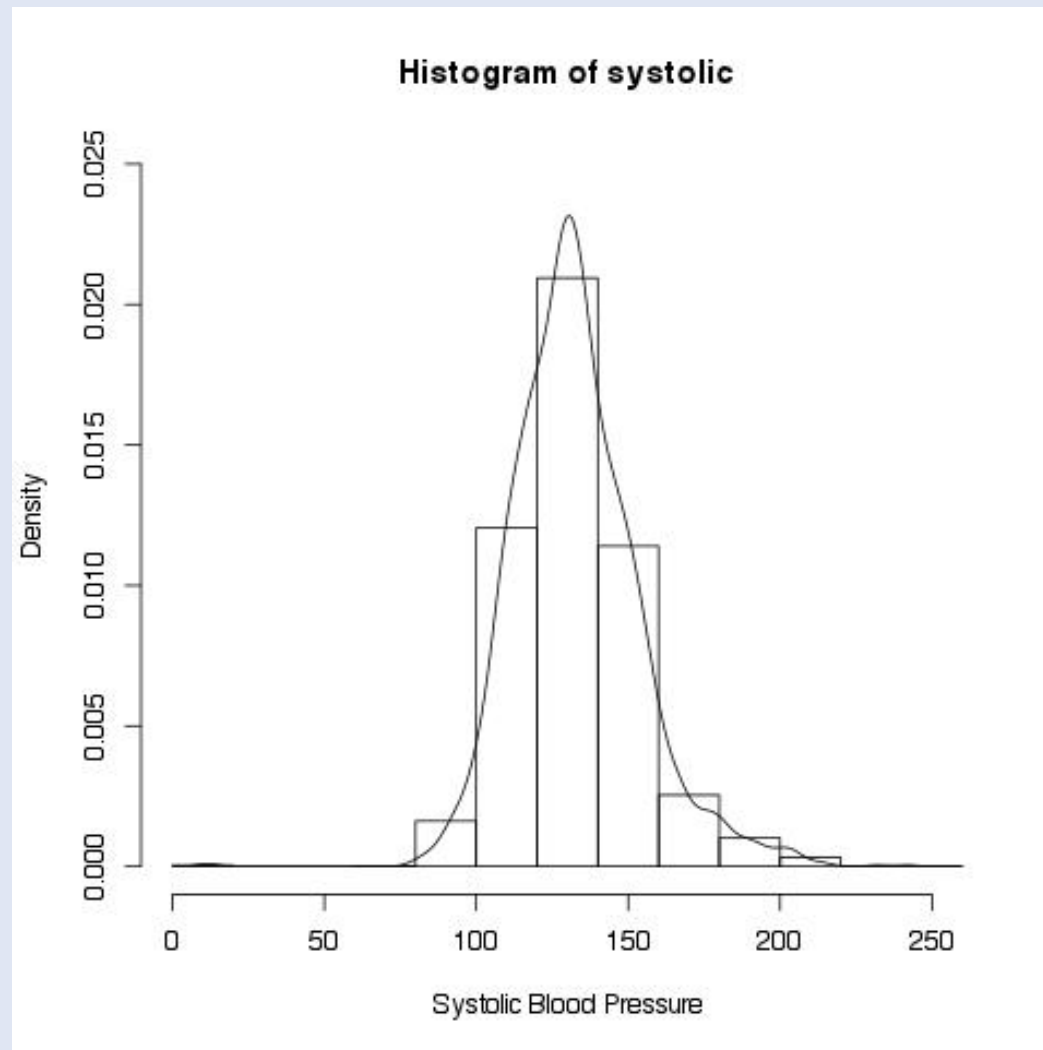
Scatter Plot

Advantages

- Allows to quickly see relation between x and y variables

Disadvantages

- Difficult to show association between more than two variables
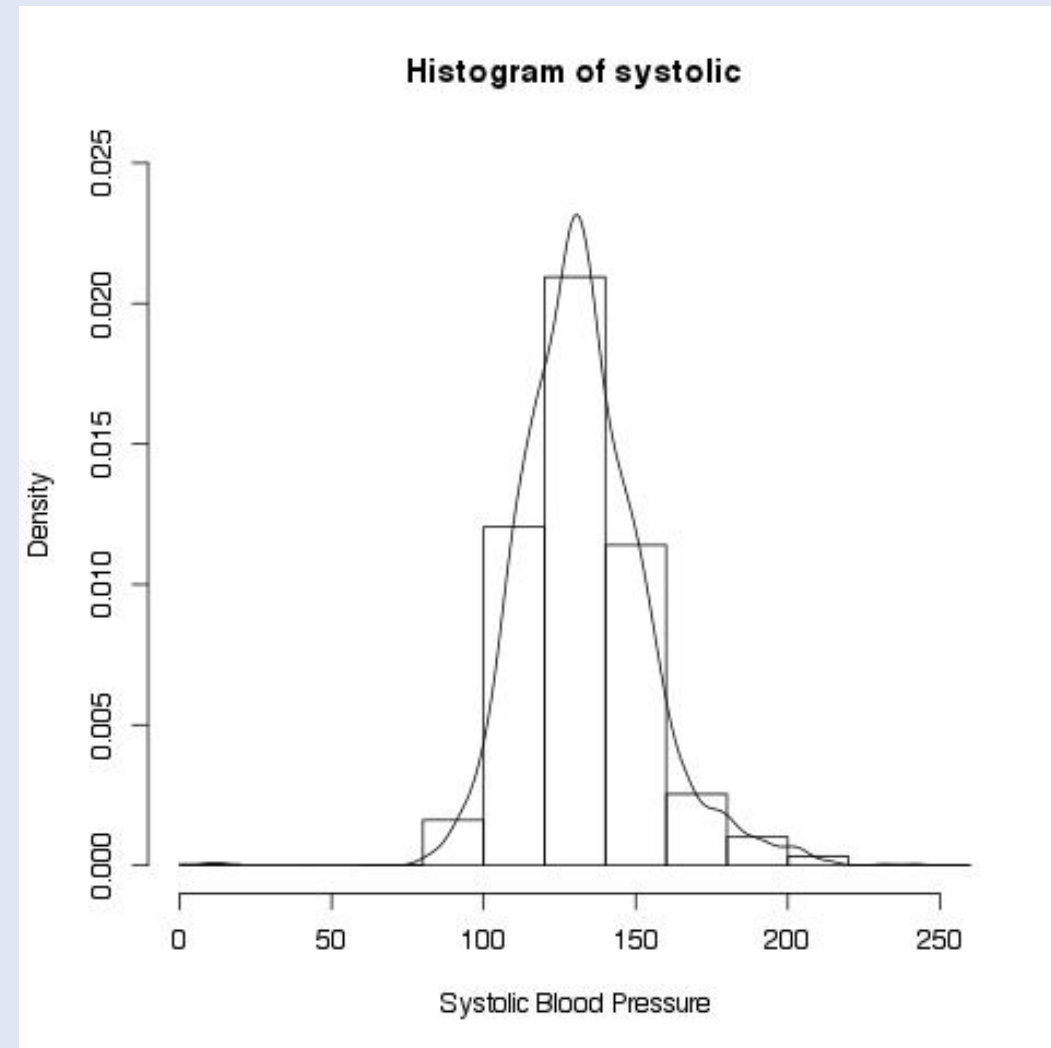
- Overwhelming with large quantity of data

3: Density Histogram

Density Histogram

Advantages

Disadvantages

Histogram of systolic

Density

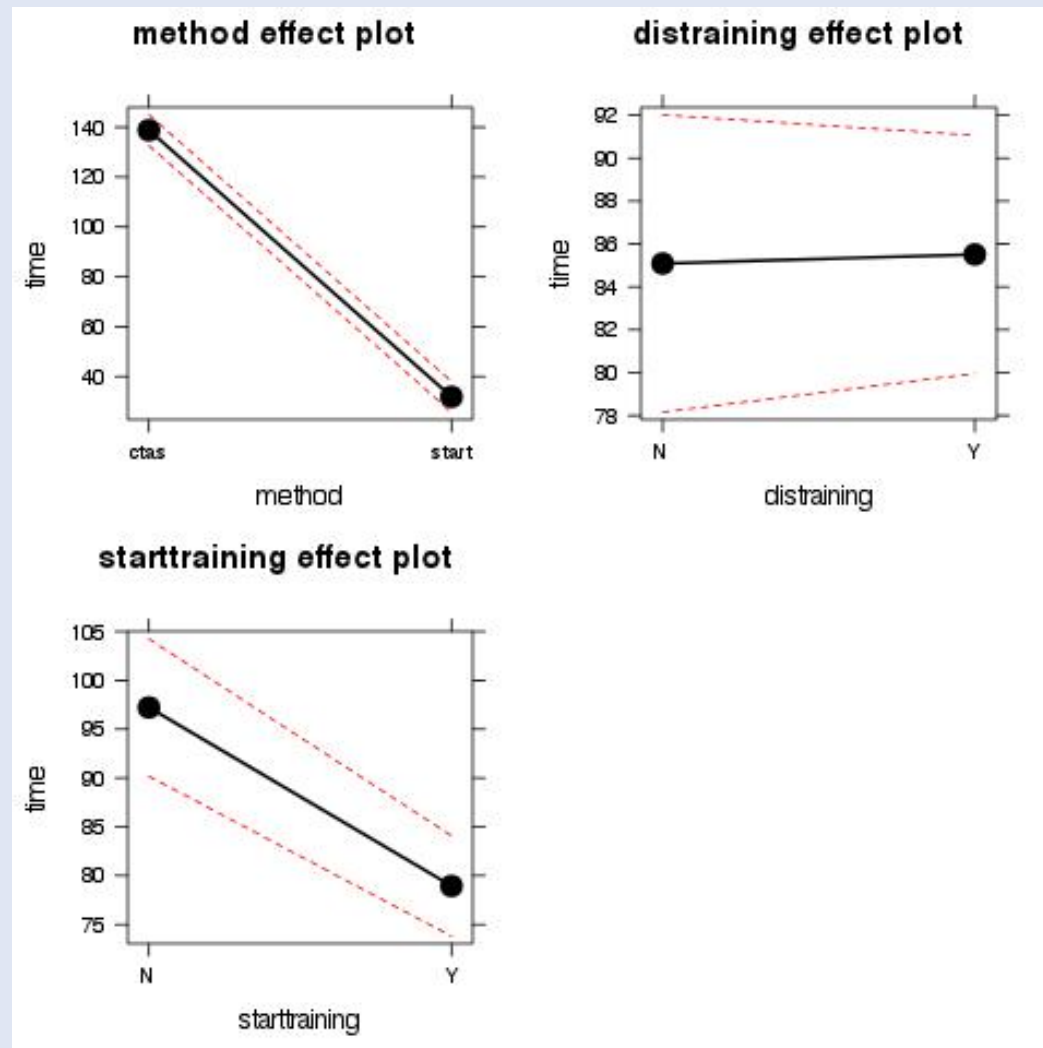Systolic Blood Pressure

Density Histogram

Advantages

- Good overall view of data

- Good if large quantity of data

Disadvantages

- Addition of density curve lose true units of measurement

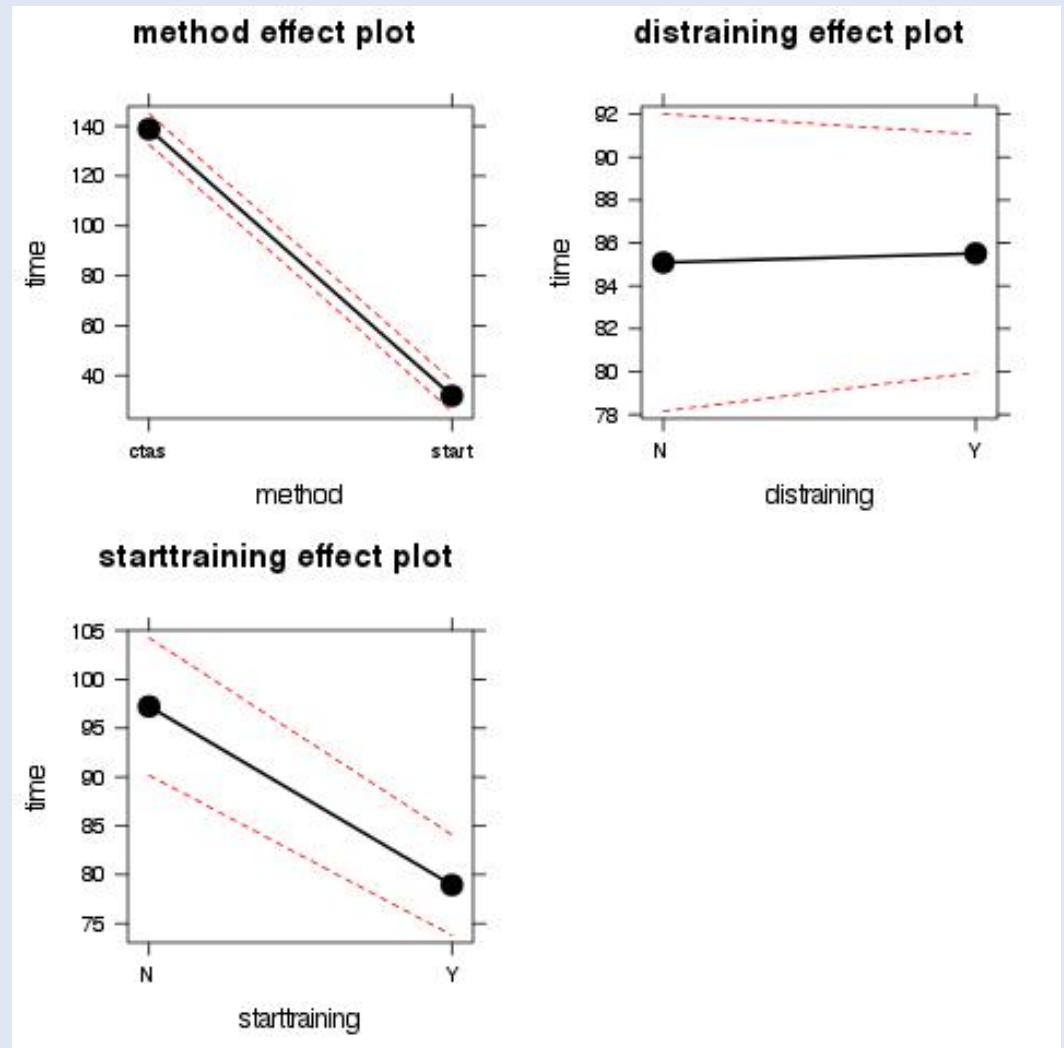- Looses fine detail of each observation

4: Effect Plots

Effect Plots

Advantages

Disadvantages

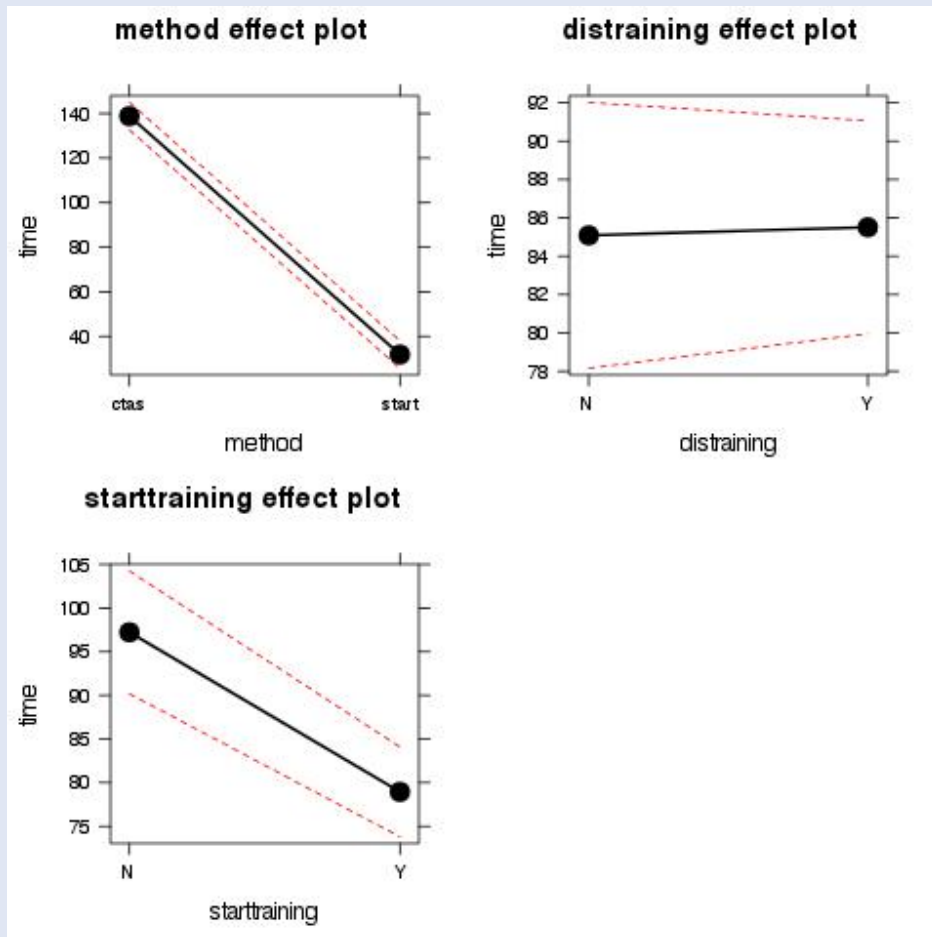# Effect Plots

Advantages

- Graphical representation of how important certain factors are

Disadvantages

- Difficult to interpret for some readers

4: Effect Plots

What is wrong with this plot?

4: Effect Plots

method effect plot

distraining effect plot

starttraining effect plot

What is wrong with this plot?
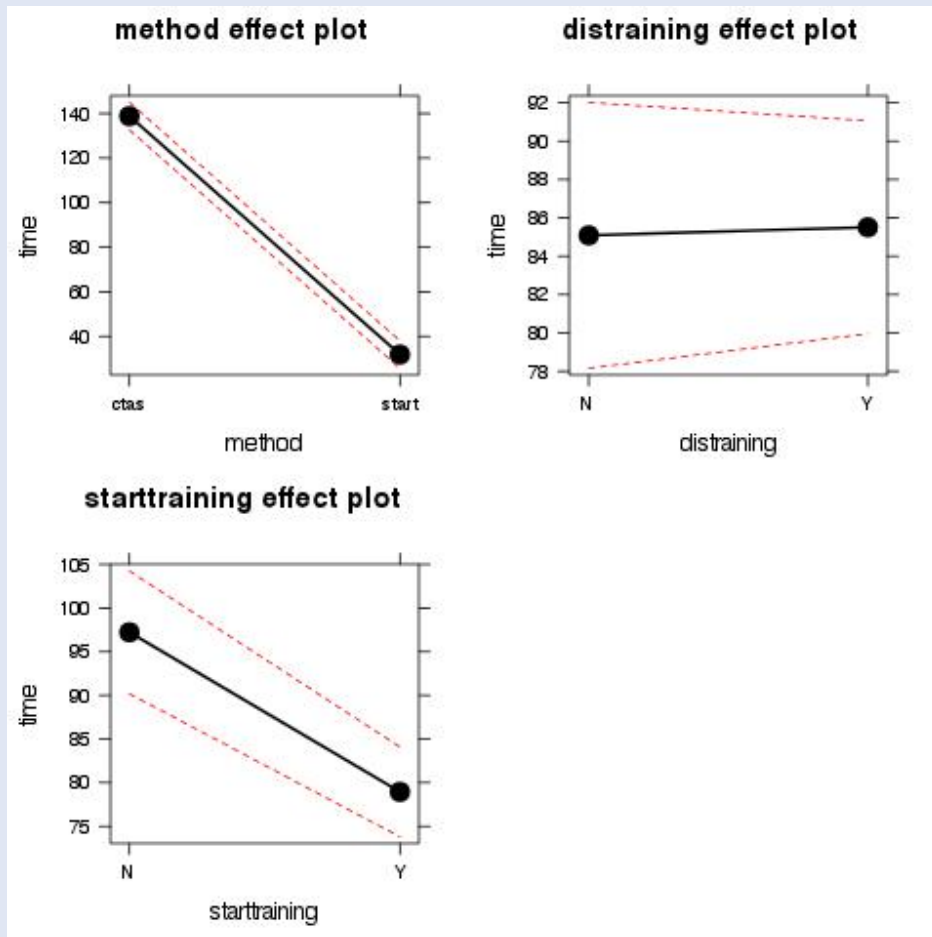
No Units

5: Barplot
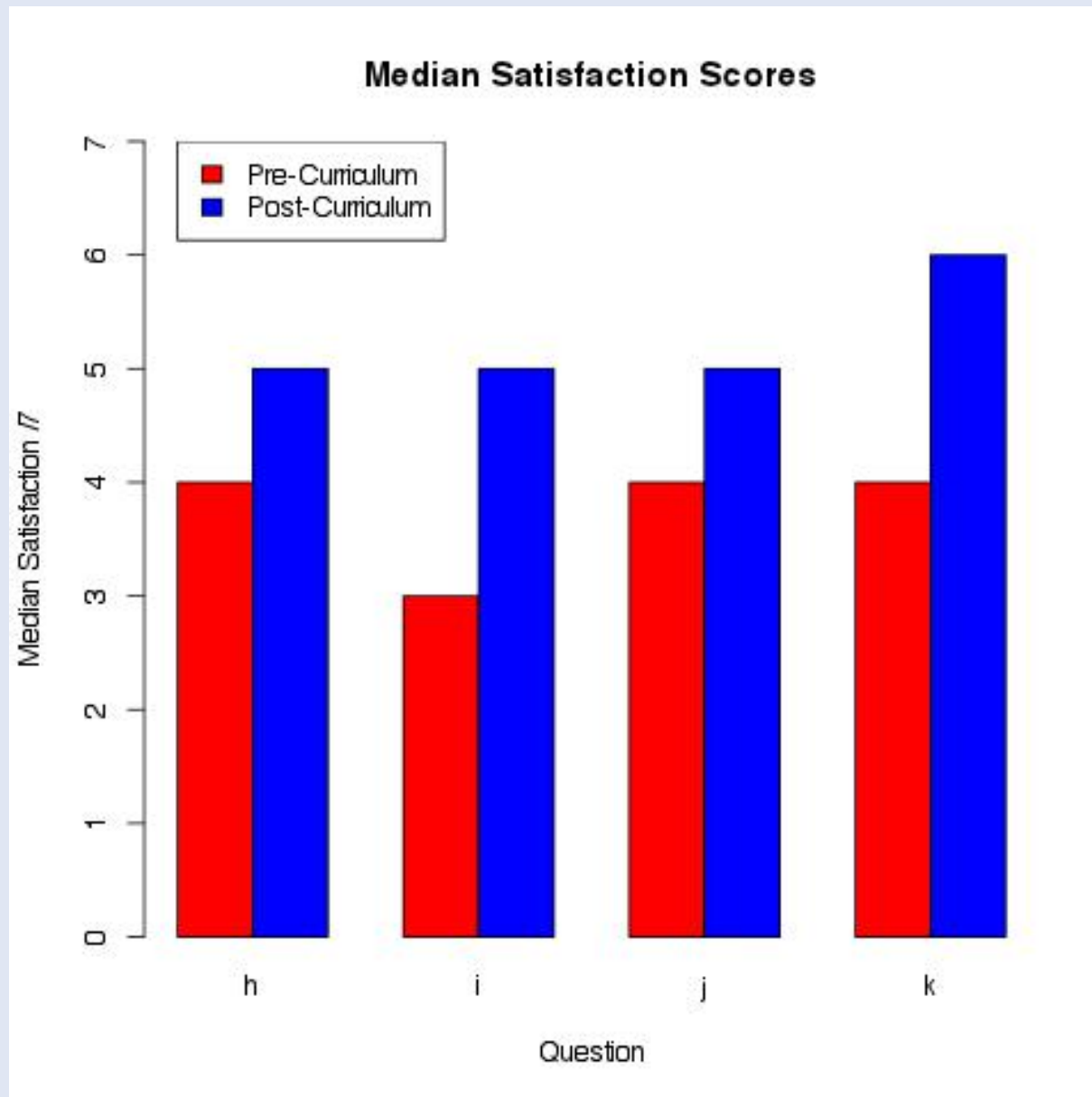
Barplot

Advantages
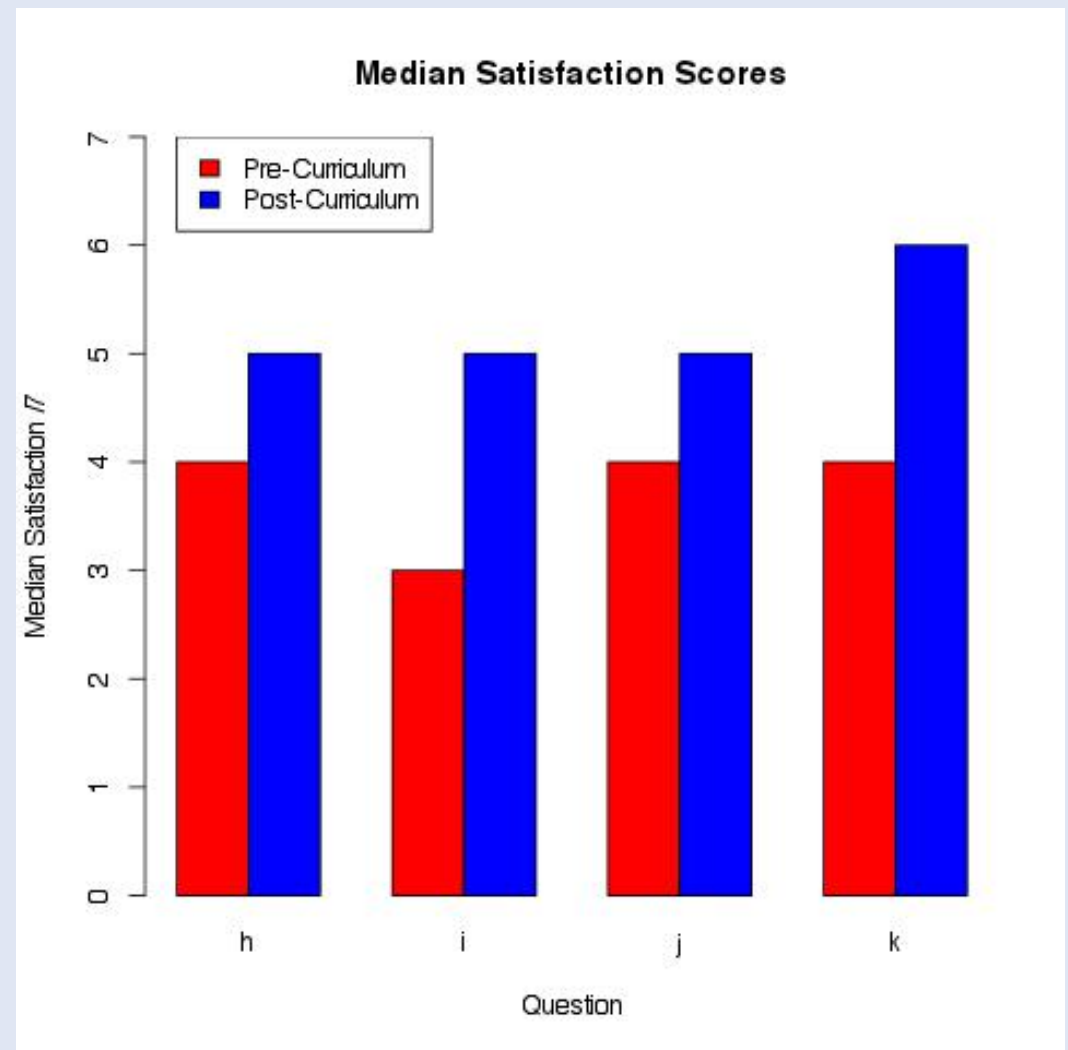
Disadvantages

Barplot

Advantages

- Small differences are easily seen

- Allows comparison of multiple groups

Disadvantages

- Difficult to interpret as number of levels of factors increase

5: Barplot
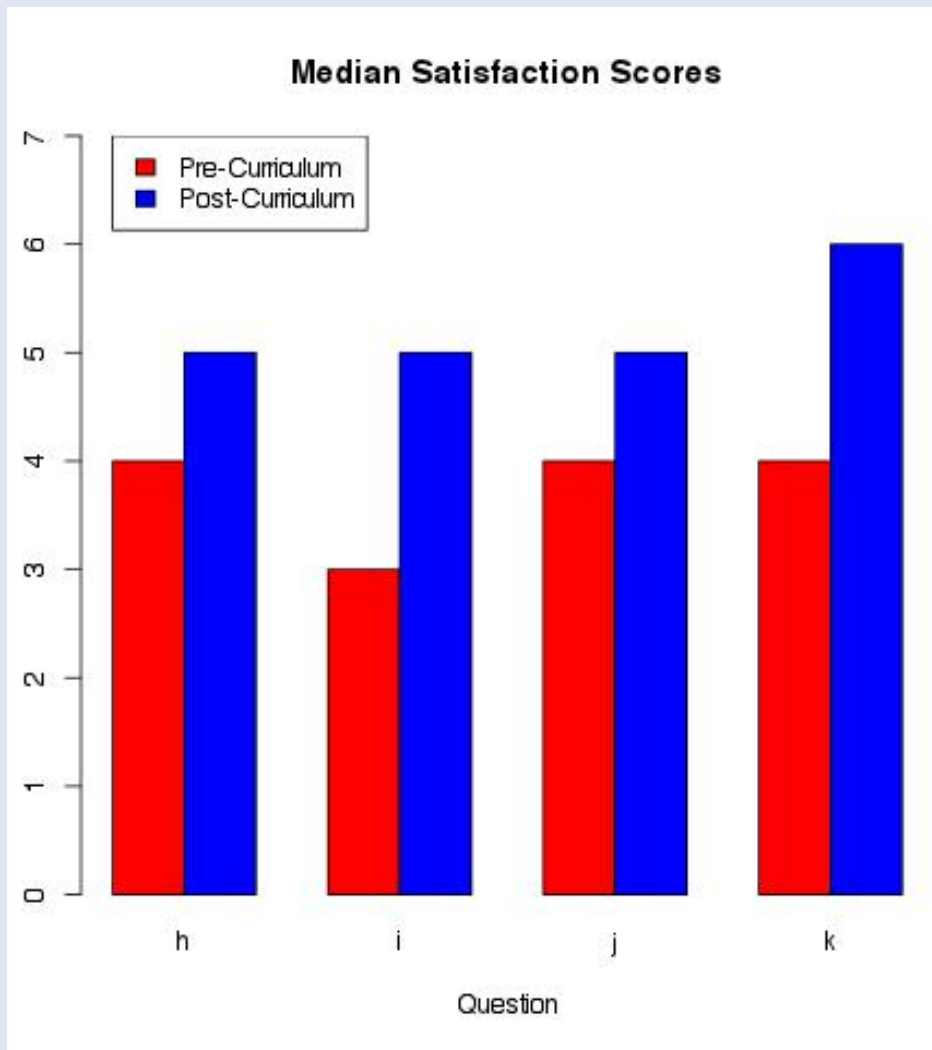
What's wrong with this Plot?

5: Barplot

Median Satisfaction Scores
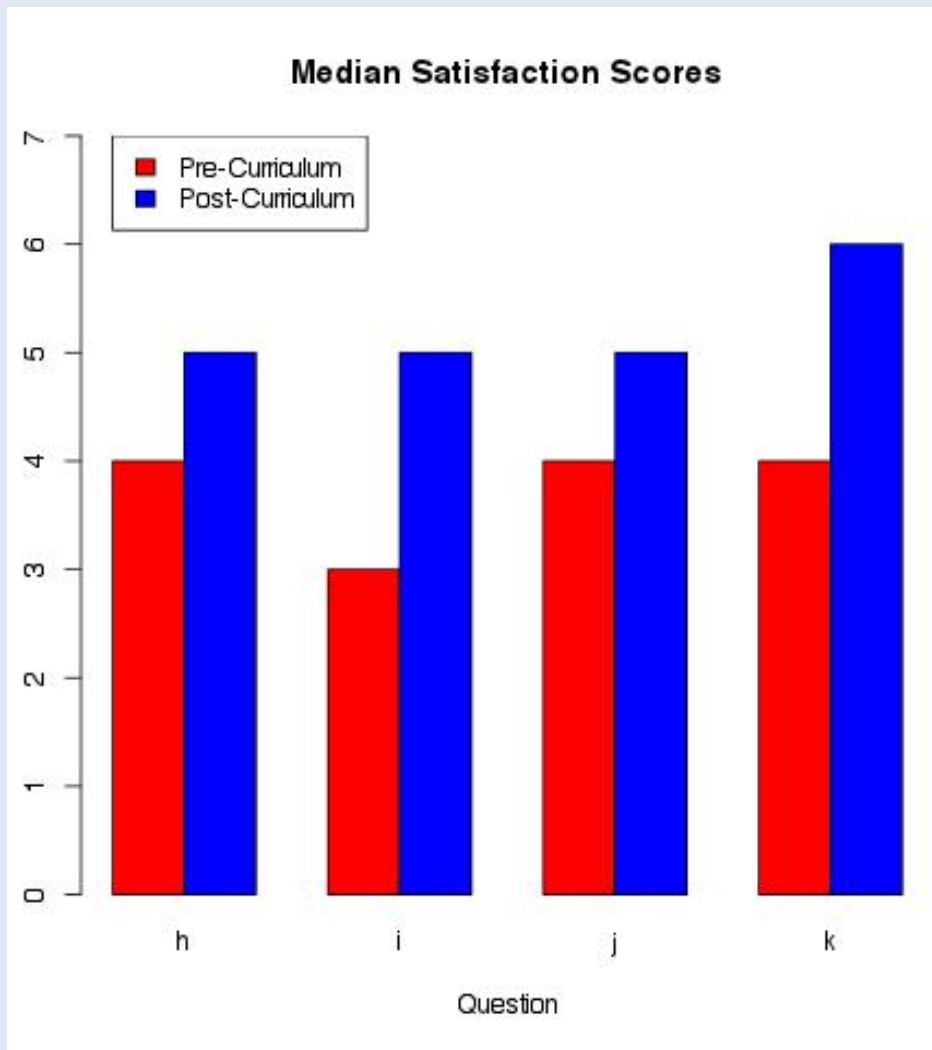
What's wrong with this Plot?

COLOR

6: Boxplot

Distribution of Triage Time by Method

Boxplot

Advantages

Disadvantages

Distribution of Triage Time by Method
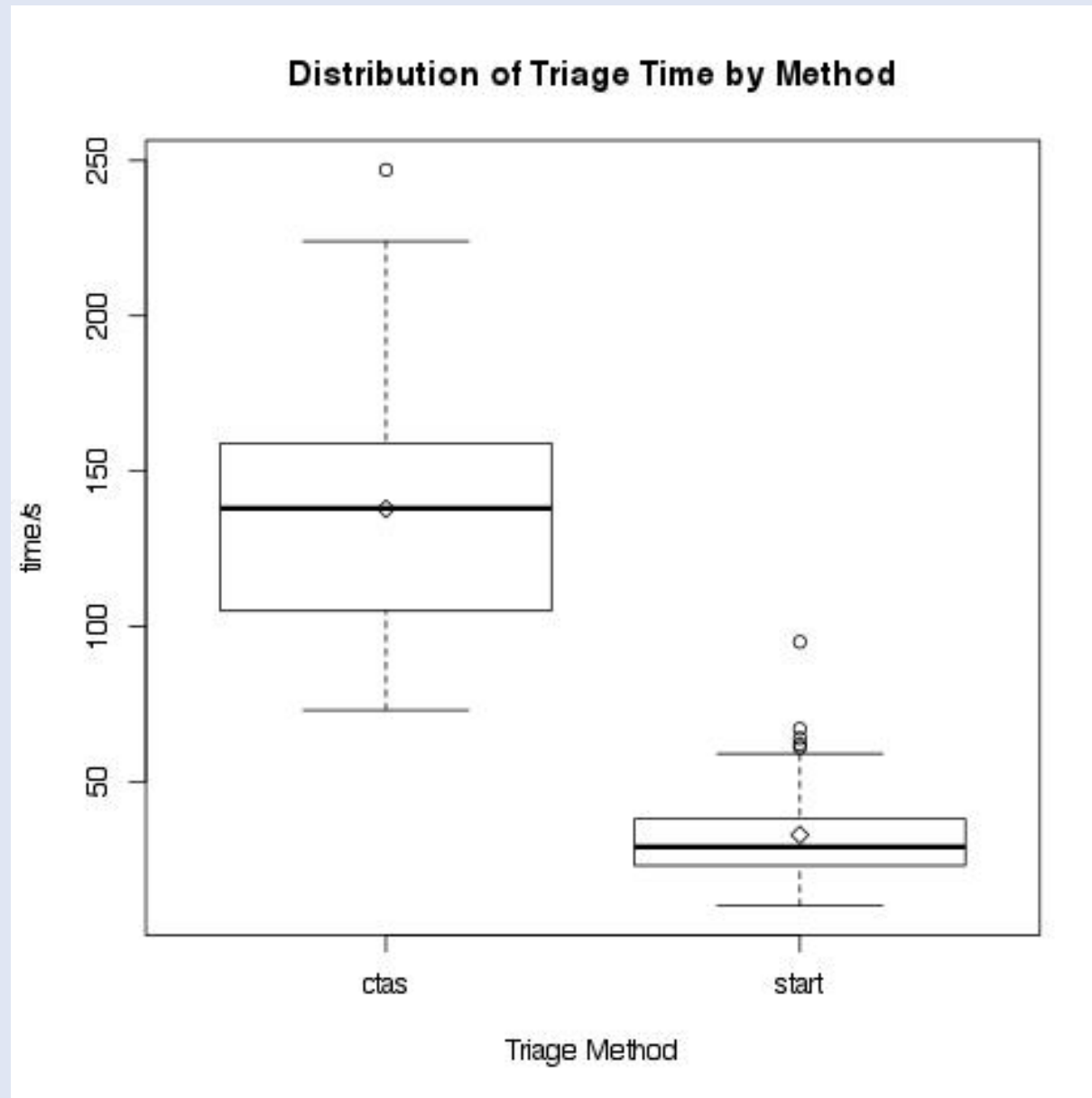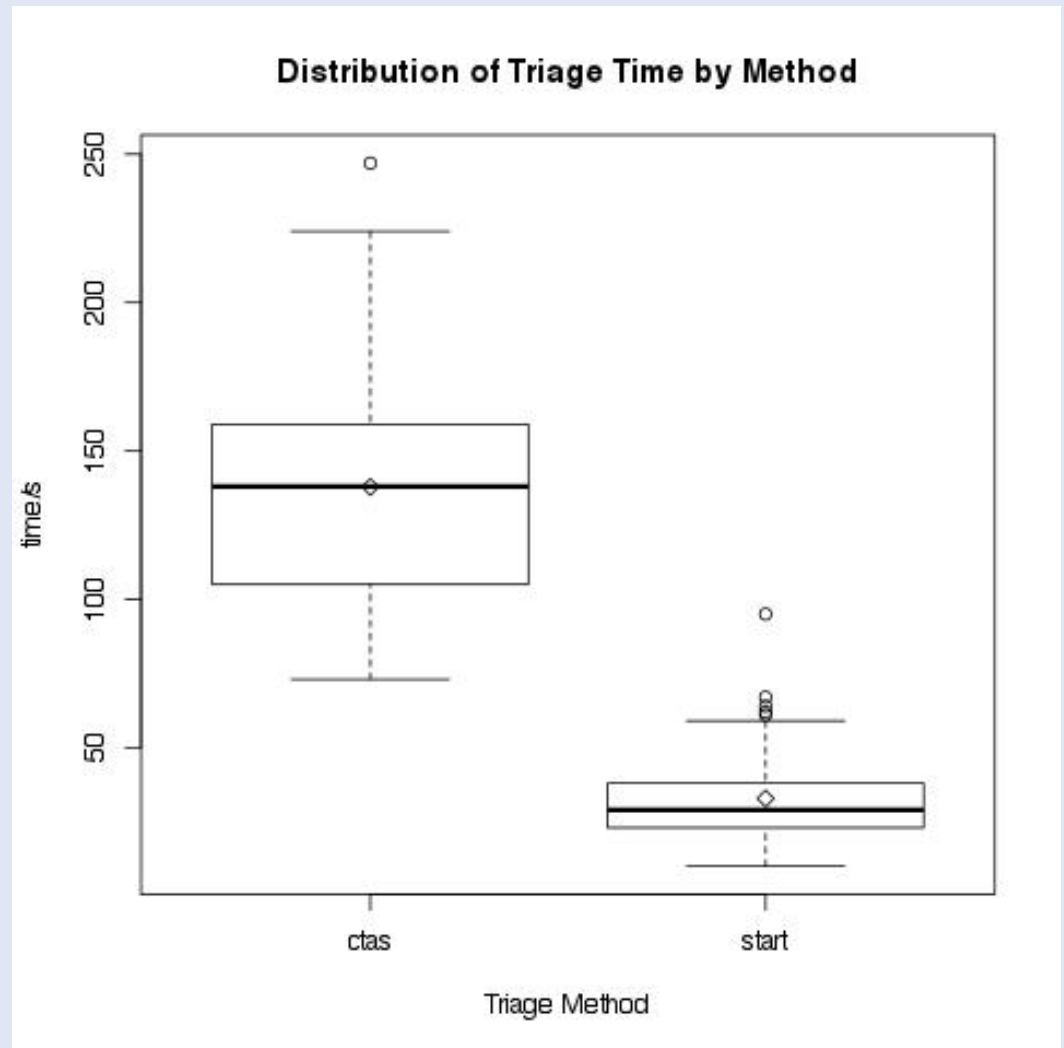
Boxplot

Advantages

- Excellent way to categorize distribution of sample

- Large amount of data in one plot

Disadvantages

- May be difficult to understand to non-statisticians

- Consider the audience

7: Pareto Chart

Pareto Chart : Proportion Incorrect by Category
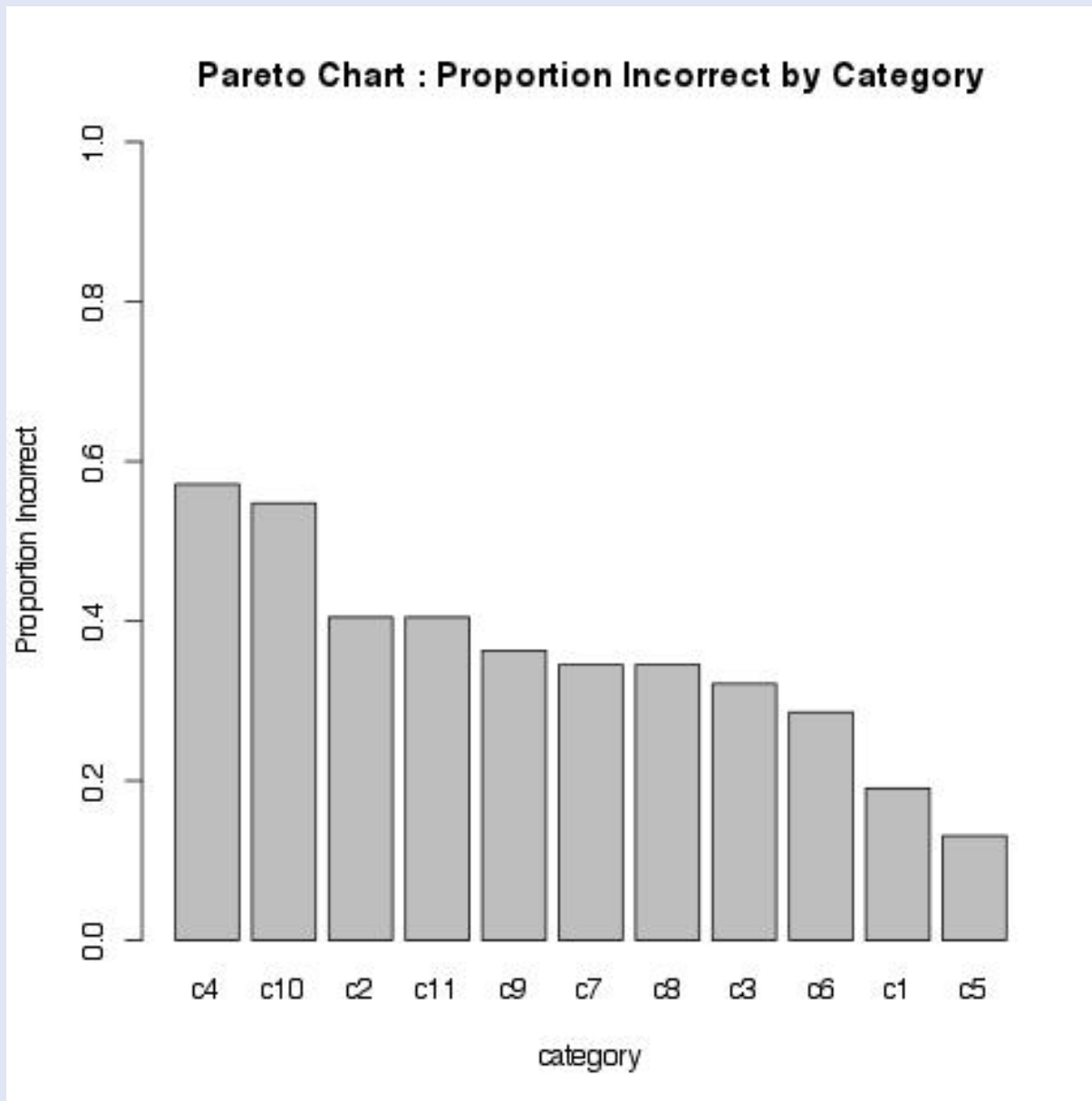
Pareto Chart

Advantages
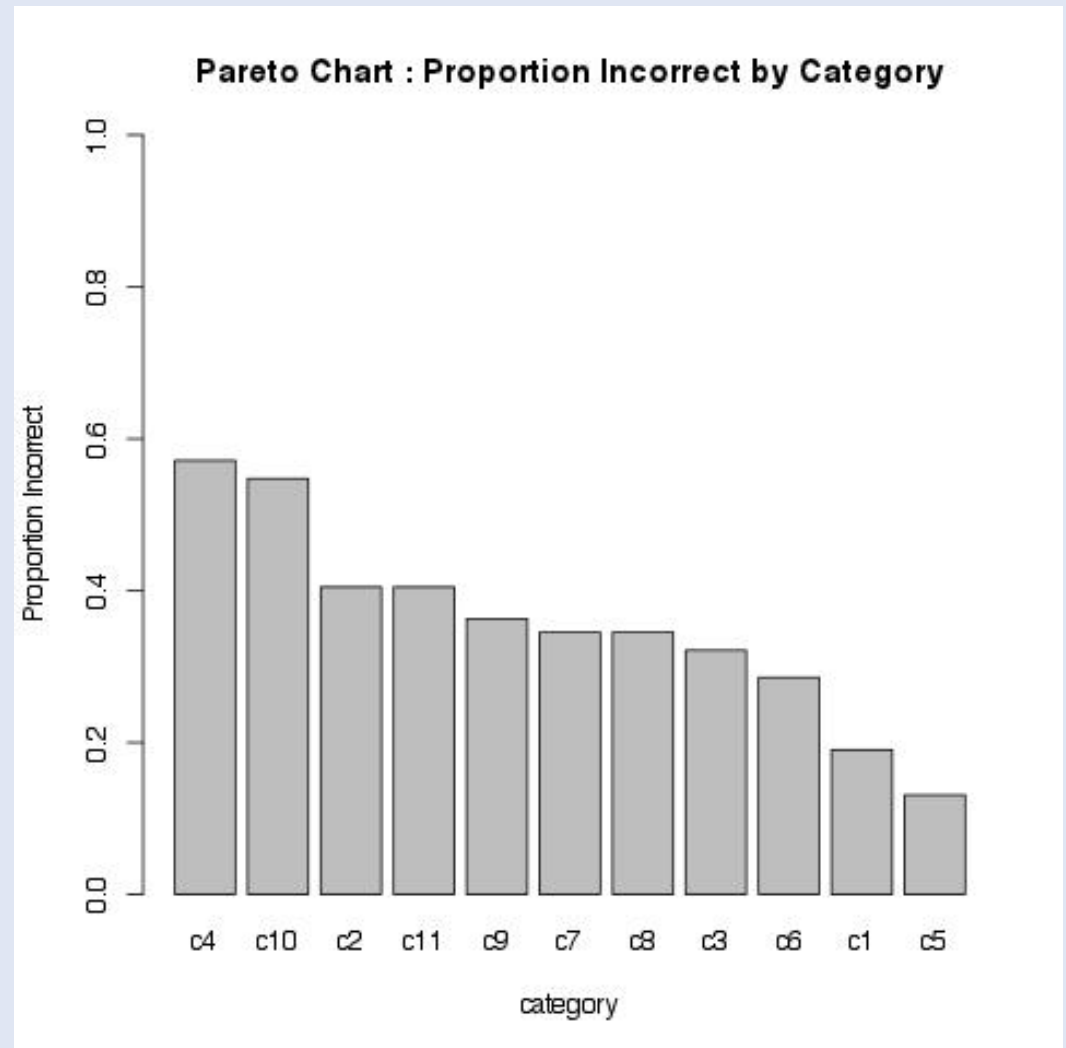
Disadvantages

Pareto Chart

Advantages

- Adds priority to the bar graph

- Easy to understand why priorities are chosen

Disadvantages

- Generally larger bars are considered more important…may need to reverse the data

7: Pareto Chart
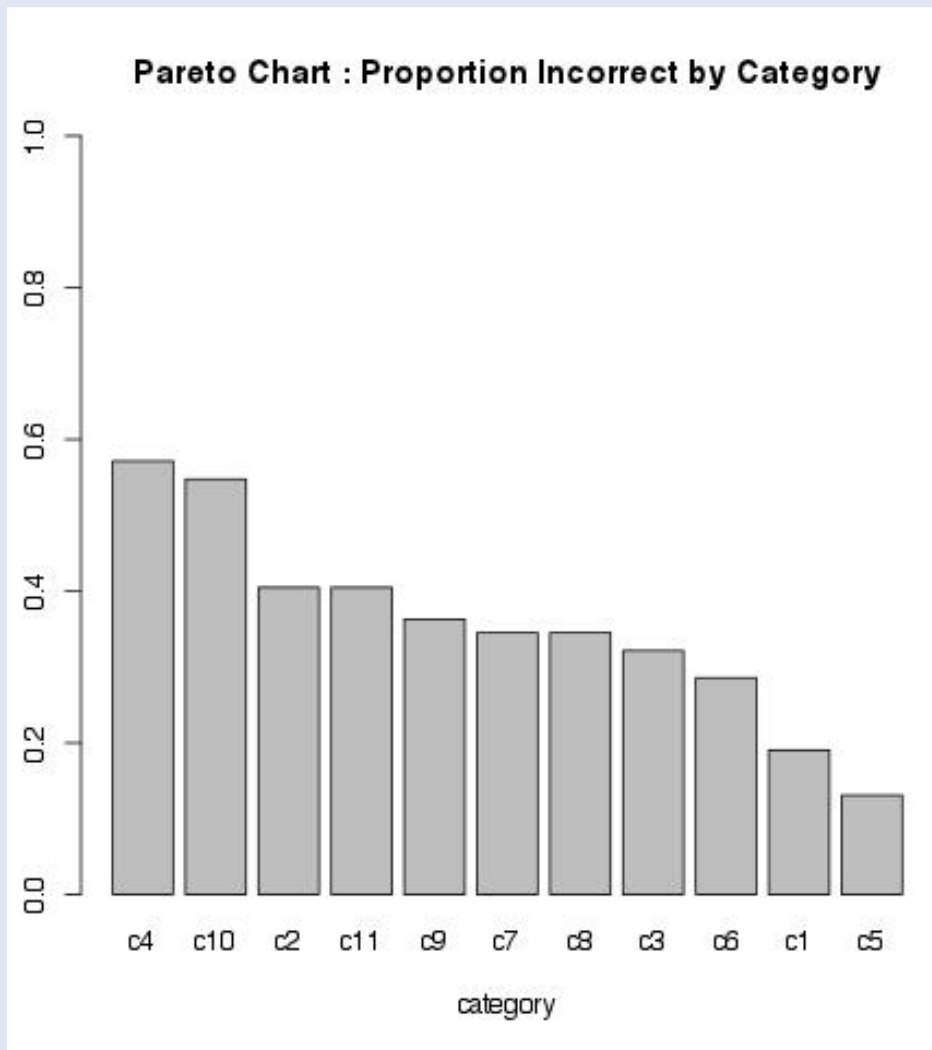
What is wrong with this plot?

7: Pareto Chart

Pareto Chart : Proportion Incorrect by Category
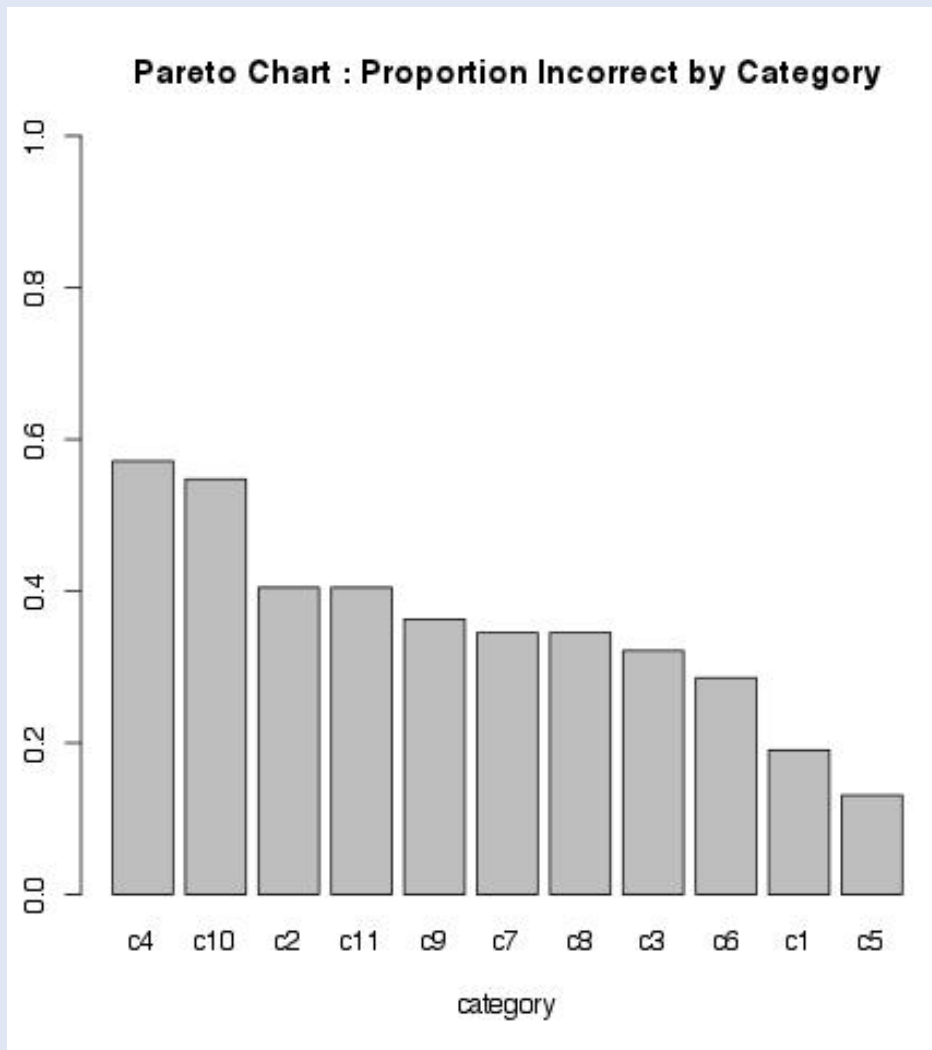
What is wrong with this plot?

No Legend

8: Control Chart

Patient Volume Per Shift

Control Chart

Advantages

Disadvantages

Patient Volume Per Shift
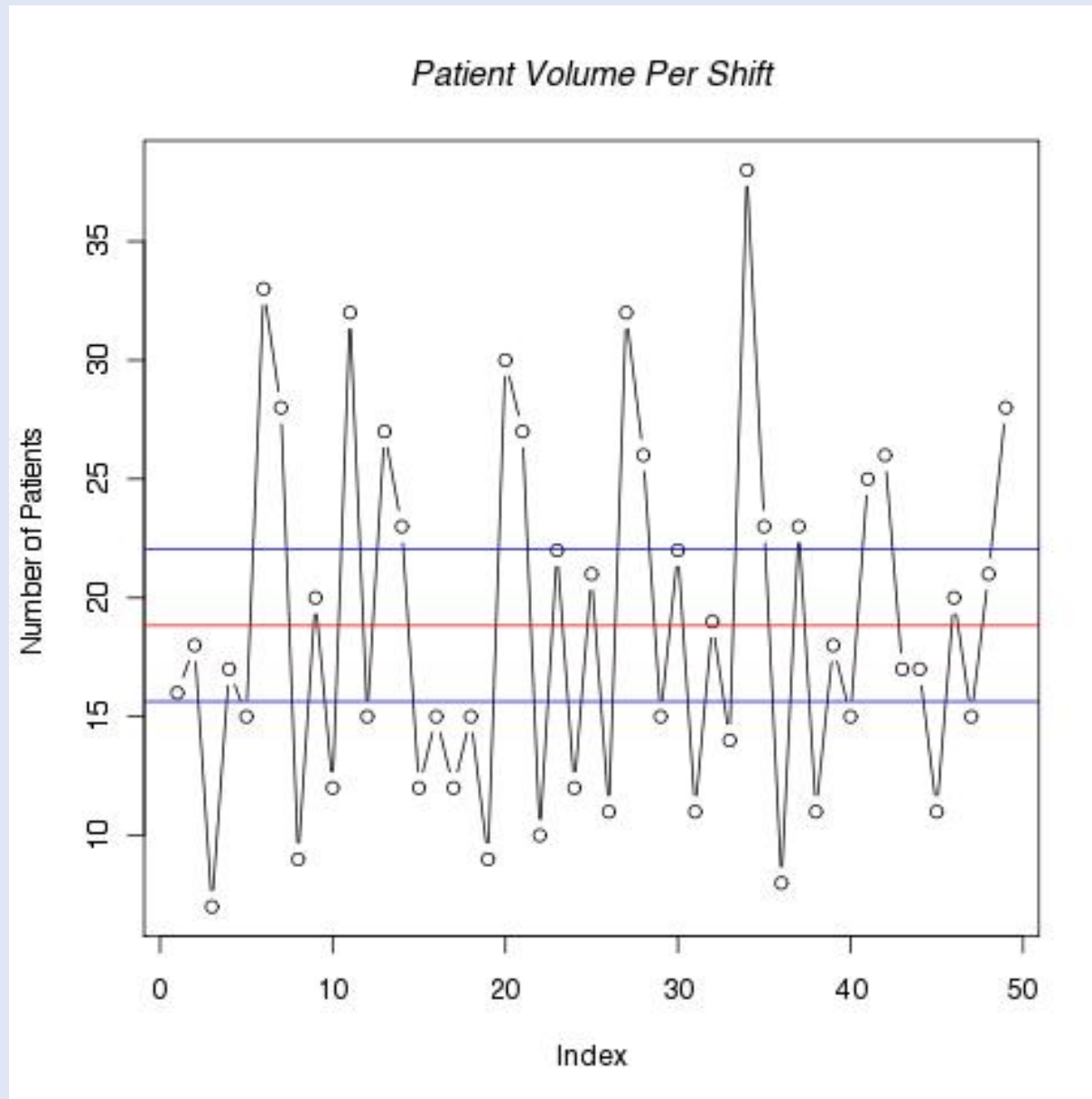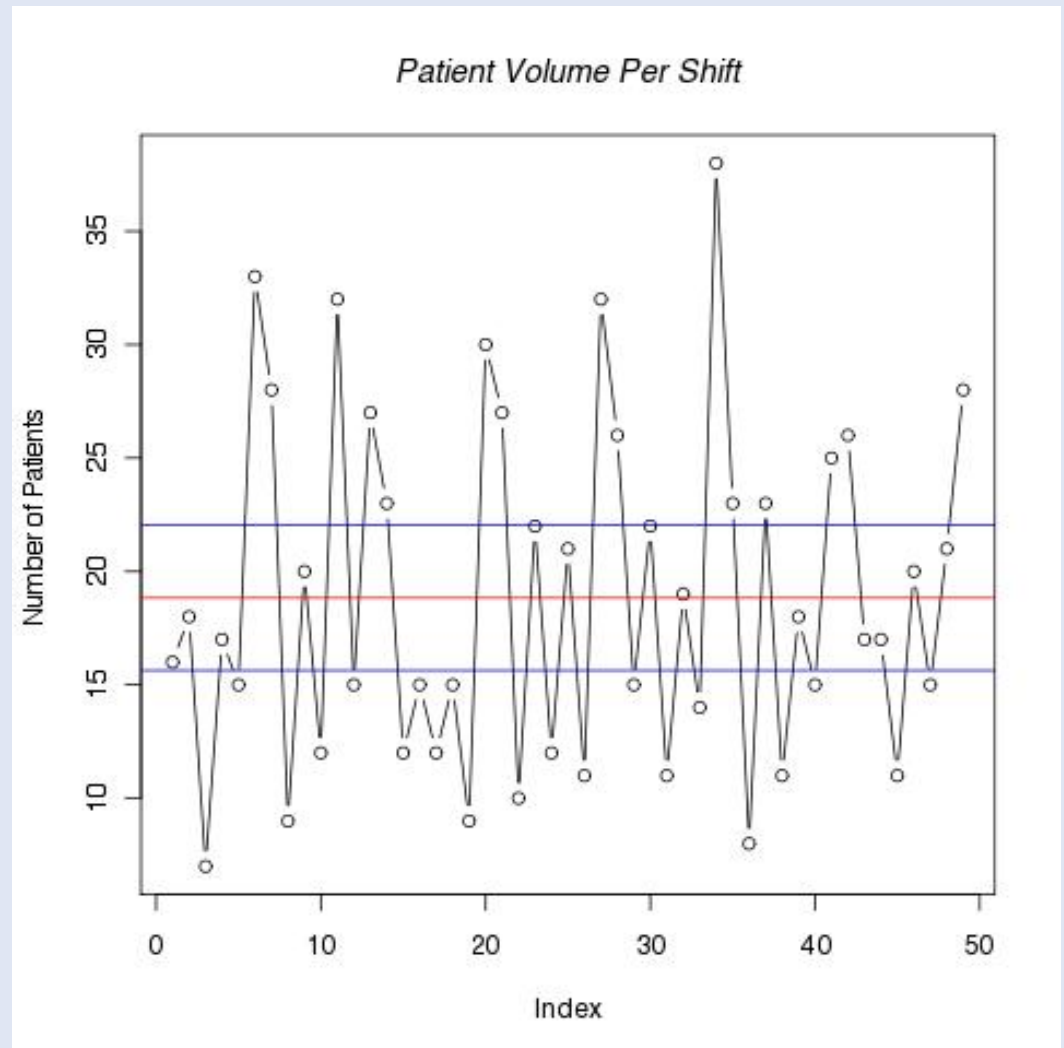
Control Chart

Advantages

- Combines statistical hypothesis testing with data visualization

Disadvantages

- Control limits often mistaken for specifications

- Need to understand principle of repeated testing

9: Analysis of Means
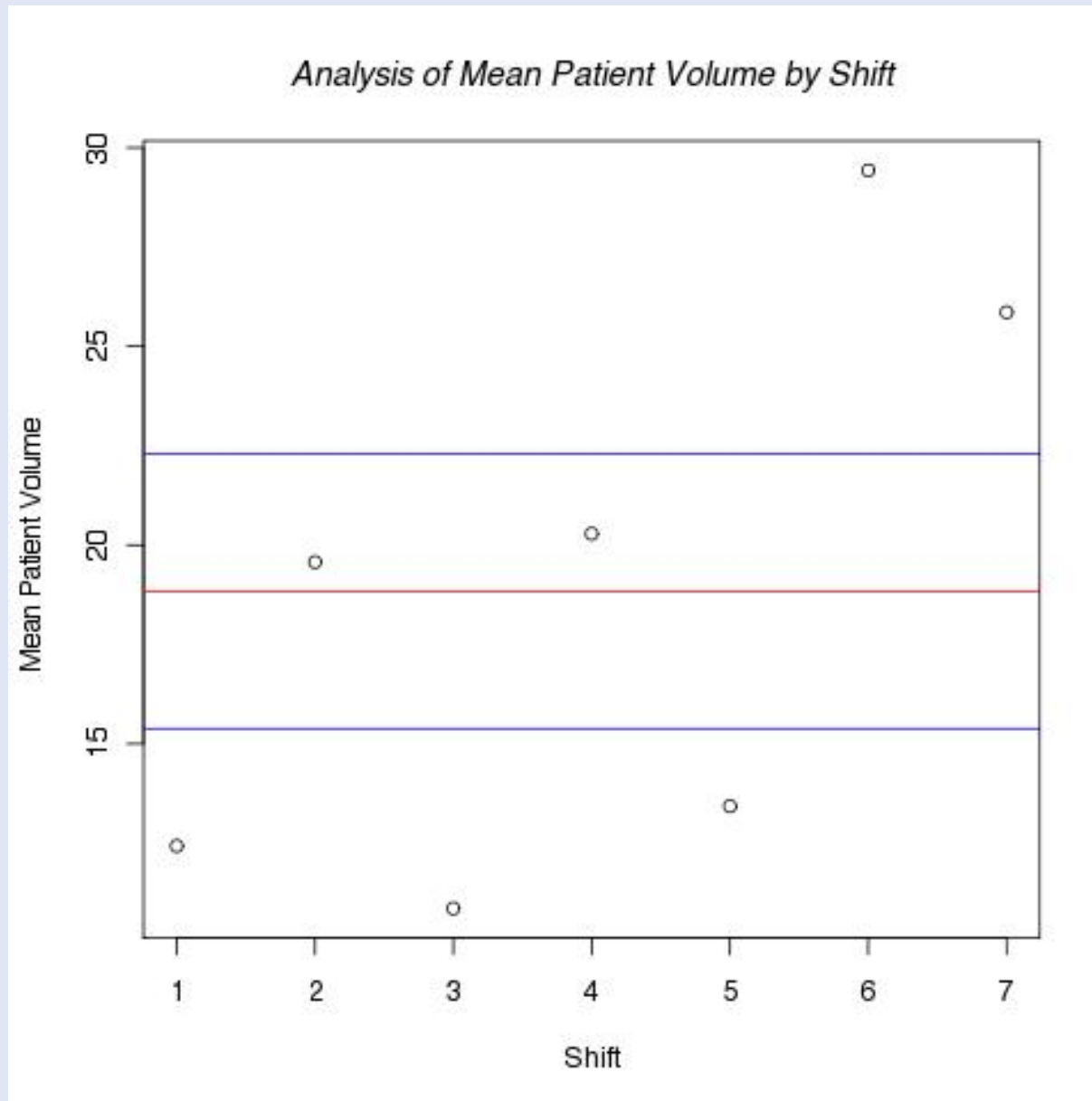
Analysis of Means

Advantages

Disadvantages

Analysis of Mean Patient Volume by Shift
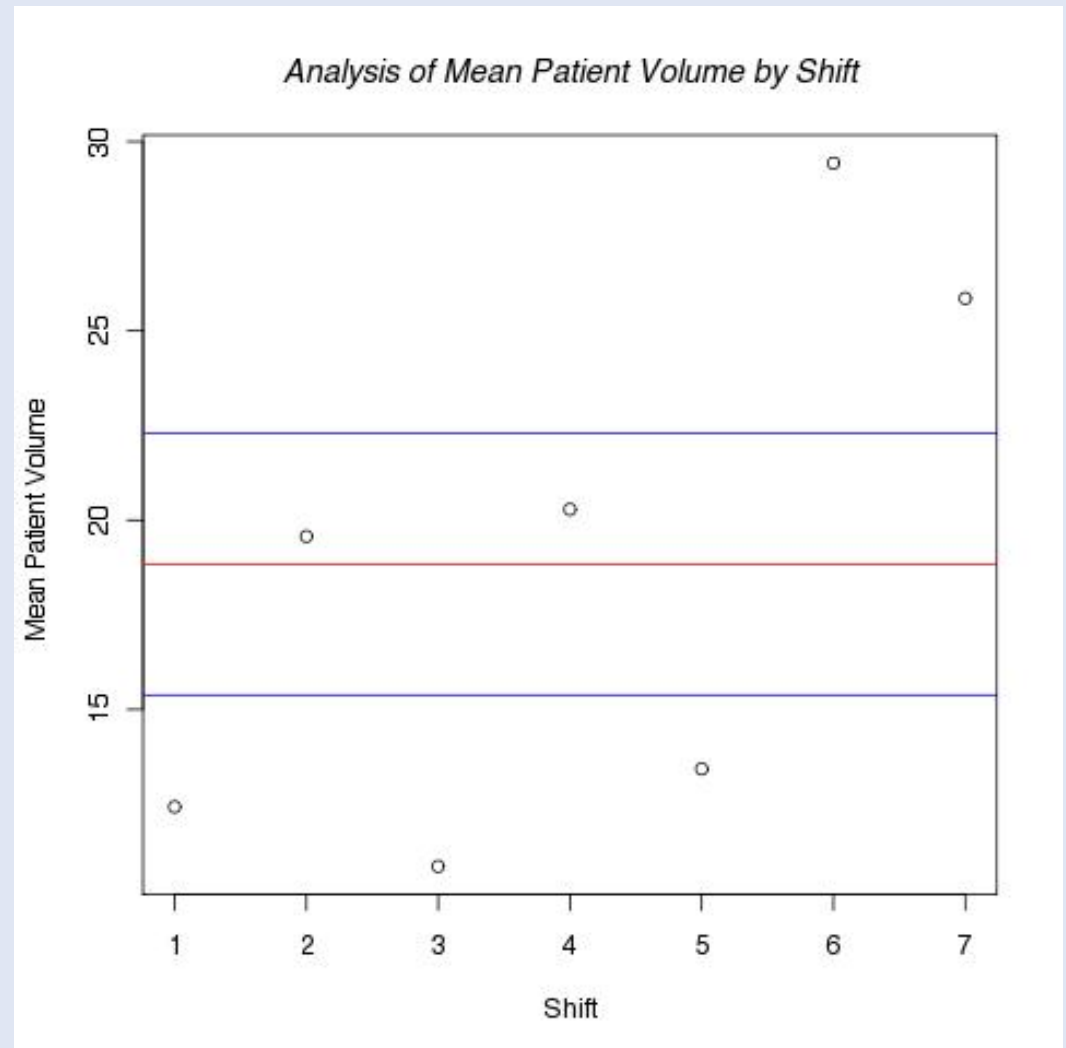
Analysis of Means

Advantages

- Graphical method to compare several groups

- Allows compensation for multiple comparisons

- Well documented

- Alternative to ANOVA

Disadvantages

- Less well known than ANOVA

- May need to explain graph to readers

10: Residual Plots
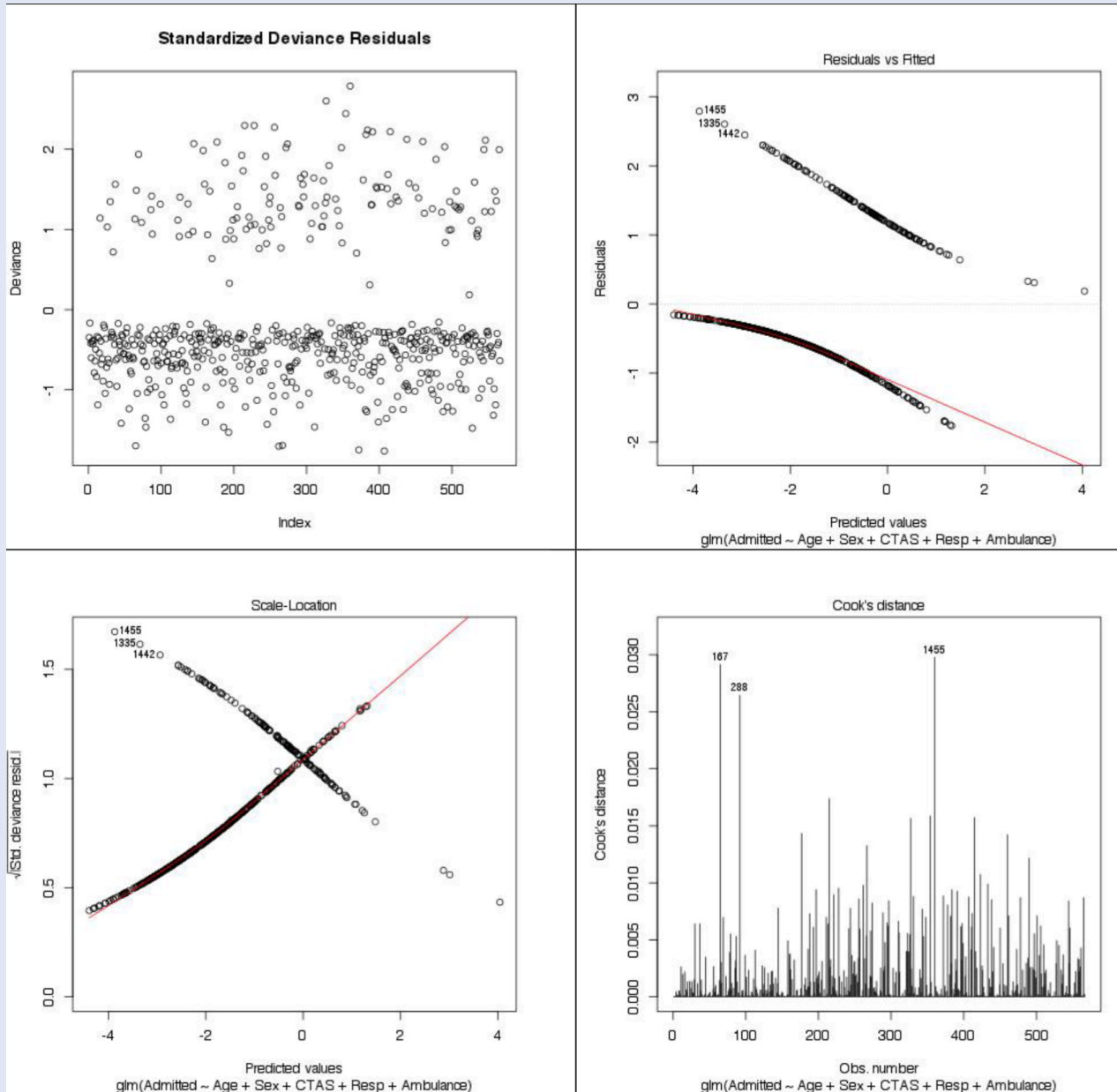
Residual Plots

Advantages
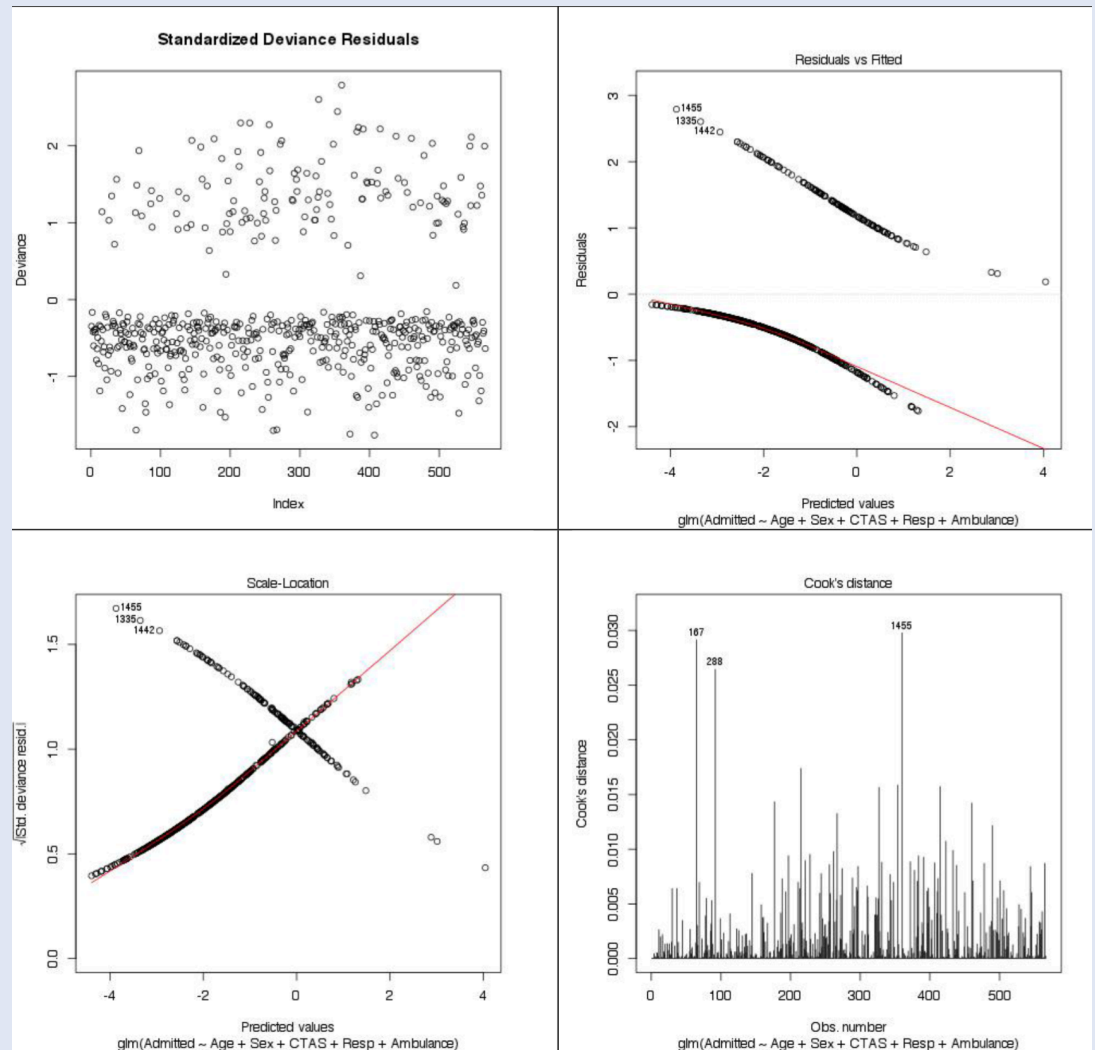
Disadvantages

Residual Plots

Advantages

- Mandatory for any study using regression

- Allow visual representation of utility of regression equation

- Useful for high-dimension multivariate data

Disadvantages

- Difficult to interpret (particularly logistic regression)

- May need explanation

(Dis)Honorable Mention: Pie Chart

Pie Chart

Advantages

Disadvantages

Education Level

Grade 6-8

Grade 1-5

Grade 9-12

Masters

Bachelors

1-4 Years College
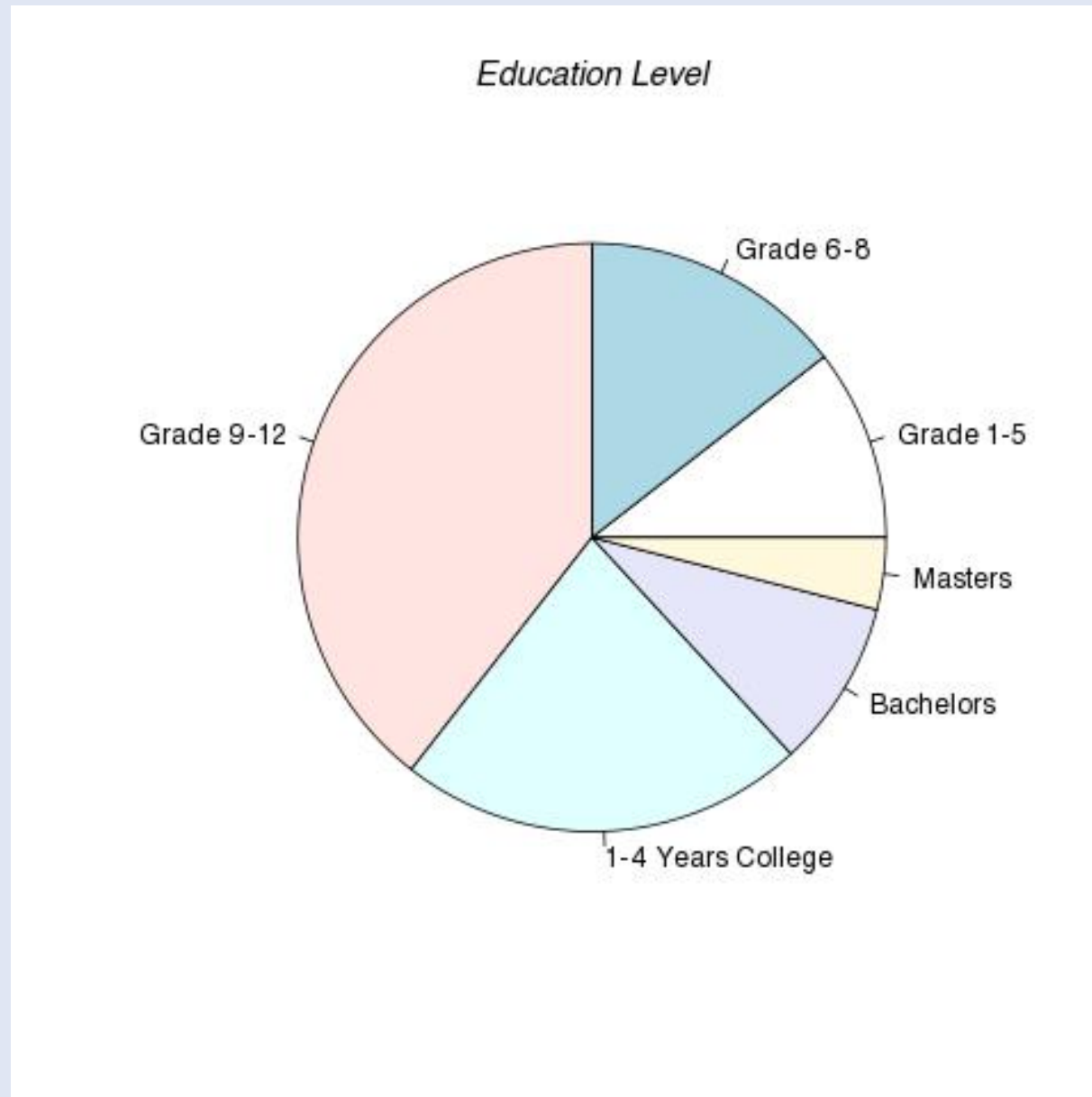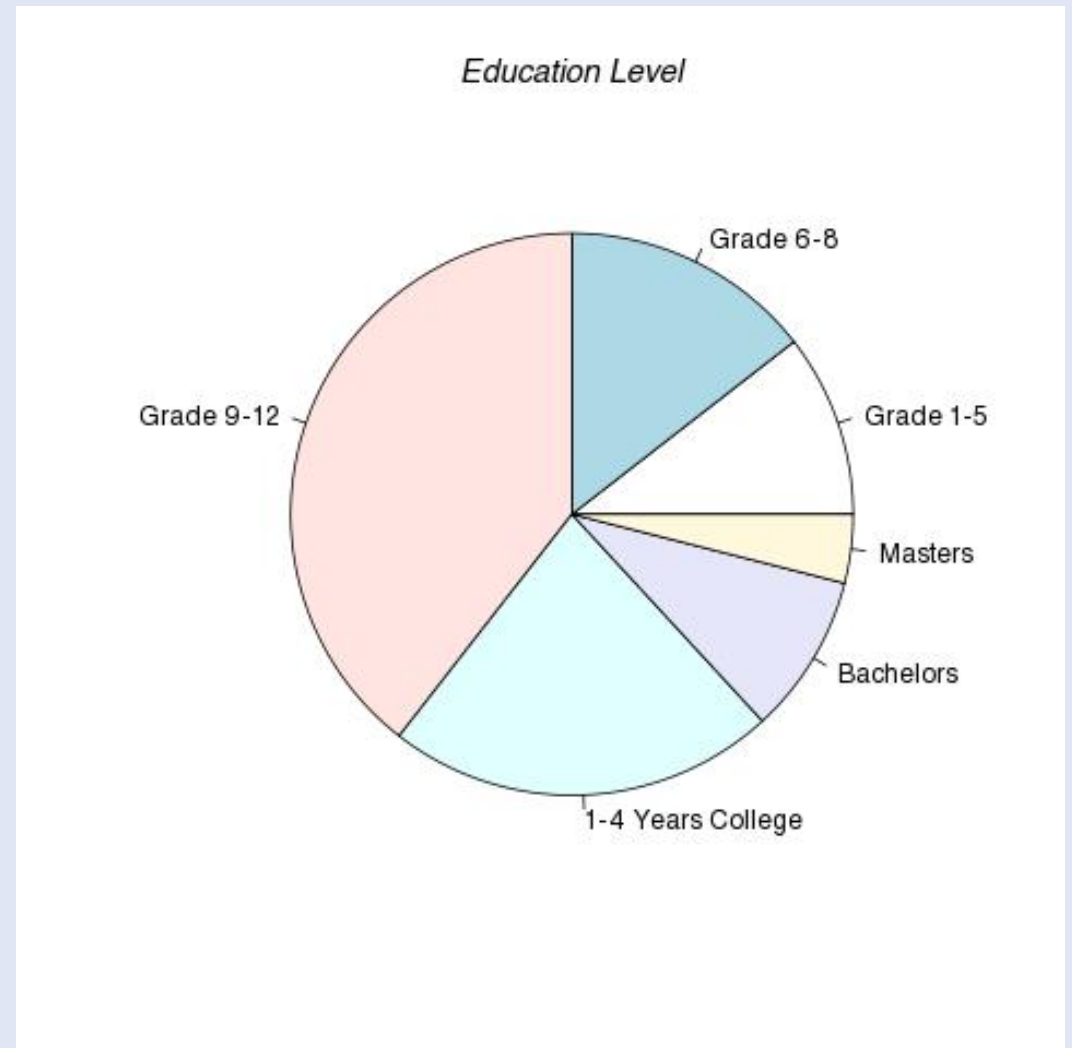
Pie Chart

Advantages

- Easy to understand for almost all readers

- Gives good overview of the composition of the whole sample

Disadvantages

- CAUTION

- Perception of area is poor compared to height

- Easy to lose subtle differences

- Not advised if close comparison is needed

- Currently low popularity rating among statisticians

Questions?

Plots

Constructing Tables

The Problem of Numeracy

A.S.C. EHRENBERG*

Lack of numeracy is due mainly to the way data are presented. Most tables of data can be improved by following a few simple rules, such as drastic rounding, ordering the rows of a table by size, and giving a brief verbal summary of the data.

KEY WORDS: Numeracy; Rounding; Ordering by size; Table layout; Short-term memory.

work; suggest why our mental processes require such rules; and consider problems of implementation.

2. TWO EXAMPLES

I start in Figure 1 with some sales statistics for eight cities in the United Kingdom. At first glance the table in Figure 1 may seem reasonably well laid out. But our attention has probably centered only on the captions—Product X; Bolton, Edinburgh, and Hull; Quarters 1 and 2; and so on. The numbers themselves are not as easy to take in. What are their main features? How can they be summarized? How can we tell someone over the phone?

Looked at with these questions in mind, the table now seems more of a jumble. It looks as if whoever produced it either did not know what the data were saying, or was not letting on. The main difficulty is that the cities are listed alphabetically, as in a directory. There is no apparent pattern in each column.

Figure 2 therefore orders the cities by the size of their adult populations, which helps dramatically. It also uses rounding off, marginal averages, and more compact layout.

Now we can see a major pattern: the bigger the cities, the higher the sales! Exceptions are also clear, like Leeds being relatively high and Luton relatively low (averages of 270 and 25).

Trends over time are also easier to take in. Although not typical, the column averages help us see that sales in each city were mostly steady quarter by quarter, but low in QIII and high in QIV. We can also see that the QIV increases were largest in Leeds and Edinburgh.

These patterns and subpatterns are easy to see in Figure 2, especially once they have been pointed out. But in Figure 1 they are still not very apparent. This illustrates the basic criterion of a good table. In a good

1. INTRODUCTION

People often feel inept when faced with numerical data. Many of us think that we lack numeracy, the ability to cope with numbers. The message of this article is that we are not to blame: The fault is not in ourselves, but in our data. Most data are badly presented and so the cure lies with the producers of the data.

To draw an analogy with literacy, we do not need to learn to read better, but writers need to be taught to write better. Luckily, numerical data have inherent structure. This makes numbers easier to communicate than ideas or verbal arguments. These few simple rules or guidelines can work wonders in communicating a table of numbers.

1. Giving marginal averages to provide a visual focus;

2. Ordering the rows or columns of the table by the marginal averages or some other measure of size (keeping to the same order if there are many similar tables);

3. Putting figures to be compared into columns rather than rows (with larger numbers on top if possible);

4. Rounding to two effective digits;

Tables

Rules for constructing tables:
1. Marginal Averages
2. Ordering of rows and columns
3. Comparative figures in columns
4. Round to 2 effective digits
5. Use layout to guide eye
6. Brief verbal description

Tables

Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Table 1 — Time from Arrival to Triage for CTAS 1 to 5 (Rounded and with Means)

Tables

1. Give marginal averages to provide a visual focus

Tables

1. Give marginal averages to provide a visual focus

Table 1			Time from Arrival to Triage for CTAS 1 to 5 (Rounded and with Means)				
Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Tables

2. Order rows and columns by marginal averages or some other measure of size. Keep same order if many similar tables.

Table 1   Time from Arrival to Triage for CTAS 1 to 5
(Rounded and with Means)

Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Tables

3. Put figures to be compared into columns rather than rows. Larger numbers on top if possible

Table 1 — Time from Arrival to Triage for CTAS 1 to 5 (Rounded and with Means)

Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Tables

4. Round to 2 effective digits

Table 1 — Time from Arrival to Triage for CTAS 1 to 5 (Rounded and with Means)

Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Tables

5. Use layout to guide the eye and facilitate comparisons

Table 1	Time from Arrival to Triage for CTAS 1 to 5 (Rounded and with Means)					
Time/sec	1	2	3	4	5	Mean
Gothenburg	45	49	100	120	140	90
Novara	51	57	110	130	180	110
Edmonton	62	63	110	140	190	110
Calgary	75	78	90	160	160	110
Ferrara	67	90	110	180	170	120
Geneva	59	79	150	220	300	160
Mean	60	70	110	160	190	120

Tables

6. Give brief verbal summary to lead the reader to the main patterns and exceptions

"Table 1 shows that the mean time from arrival to triage was 120 seconds for all patients. Overall, the Gothenburg group triages patients faster. CTAS 1 patients were triaged more quickly than other CTAS groups."

Tables

Questions?

Objectives

- Know the advantages and disadvantages of various plots and how to choose the best plot to emphasize data

- Avoid common mistakes in plotting.

Math Lesson

Post hoc analysis of group
  differences using Bonferroni
  confidence intervals

Ouch!