Back to School

Quality 101

Being a Data Detective:

Using Data to Pinpoint Problems

Jill Finan, Process Excellence Manager/Six Sigma Master Black Belt
Ortho Clinical Diagnostics

Objectives

In this presentation, we will cover:

- Reasons for using data in problem-solving and decision-making
- Characteristics of "good" data and how to collect it
- Common data displays and what they are used for
- Practical use of data displays to solve a problem (case study)

Why Use Data?

- To separate what we THINK is happening from what is REALLY happening
- To confirm (or disprove) our preconceived notions
- To see the history of a problem over time
- To identify and understand relationships that might be causing our problems
- To avoid "solutions" that don't address the root cause and don't solve the **REAL** problem!

How to Collect Good Data

- Ask yourself this key question:

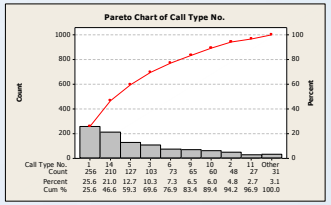
"What do I want to know?"

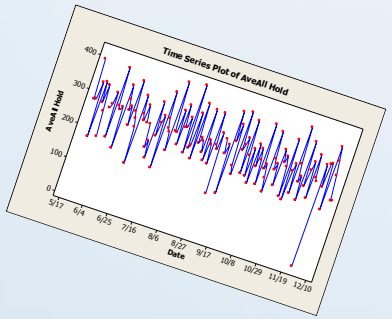
- Understand what type of data you want to collect:
  - Continuous data: needs a measurement system, can be broken down into smaller units of measure
  - Discrete data: counts, percentages, ordinals (think survey answers)
- Define the way you are going to measure; how will you ensure consistency?
- Think about the way you are going to display the data (more on this in a few!)

Data Characteristics

Useful Data are:	Typical Problem
Sufficient – There is enough data so the patterns you see are likely to be real	**Insufficient** – There is not enough data to help you draw reliable conclusions
Relevant – The data will help you understand or pinpoint the problem you are trying to solve	**Irrelevant** – The data describe a characteristic that doesn't help you understand the targeted problem
Representative – Encompasses the full range of process conditions	**Biased** – Representative of only certain process conditions
Contextual – Collected with other information about what is happening in the process	**Isolated** – The data you collect is the only information you have about the process

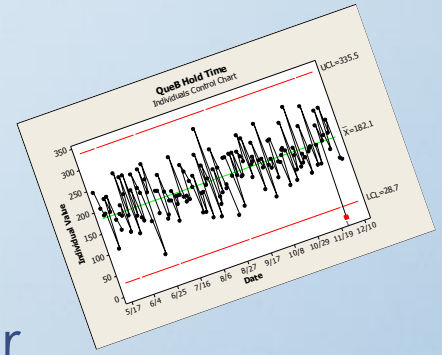
Data Displays





"A picture is worth a thousand lines of data"

- Help us track a process over time to identify patterns, problems, and opportunities
- Help us distinguish between special cause and common cause variation
- Reveal the characteristics of our data – shape, center, spread – to enable statistical analysis
- Help us break a big problem down into parts so we can focus on the most important aspects
- Help us determine relationships between two or more variables



Case Study:

The Mystery in the Data

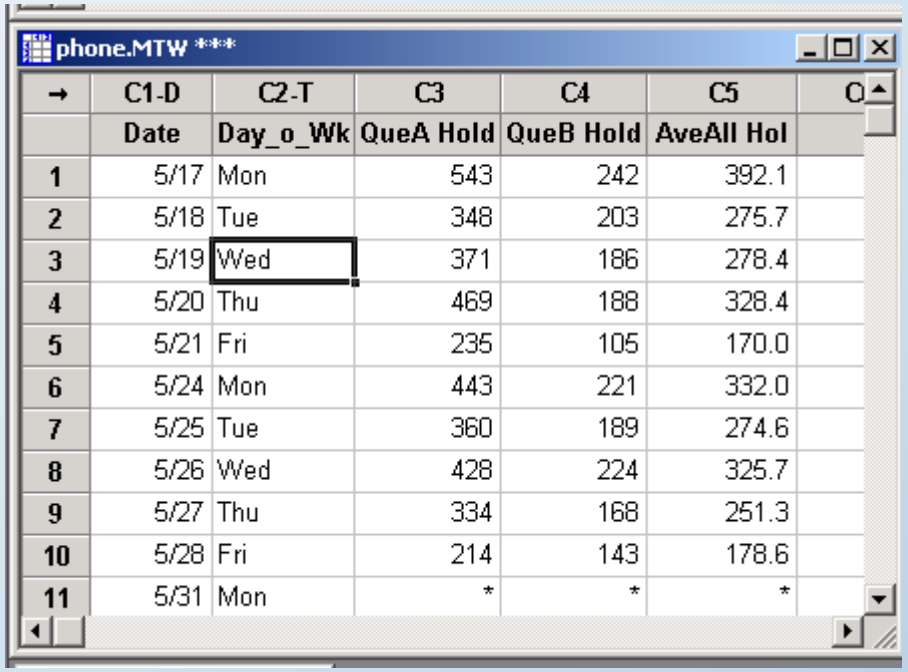
The Mystery in the Data

You work for a health insurance company. Your department staffs the call center for processing medical claims. When a customer phones the call center, a voice menu directs them to "press 1" if the claim is less than $200 or "press 2" if the claim is more than $200. They are then placed on hold in one of two queues, called "QueA" or "QueB," until a representative is available to answer the call.

- Many customers are defecting to other insurance providers because of the slow claims authorization process. A survey shows that 75% of customers perceive the hold time for a representative as too long.
- Your project team is working to decrease the hold times.
- What do you want to know? How long are customers holding!
- You obtain a history of the average daily hold times stored in the call system database.

Your Data

- You collect data over a period of 8 months, 5 days a week, for the 2 queues in the call center.
- Note the relevant data that was also collected.



phone.MTW ***

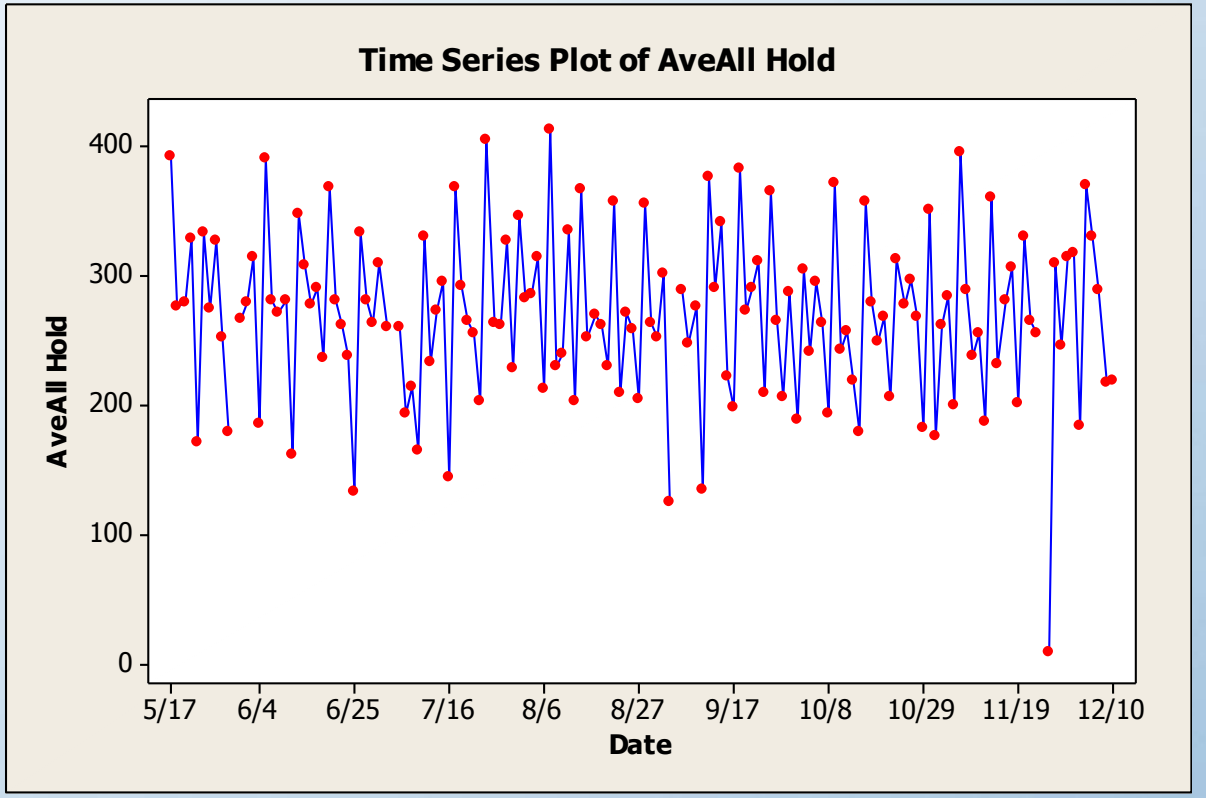
→	C1-D	C2-T	C3	C4	C5	C
	Date	Day_o_Wk	QueA Hold	QueB Hold	AveAll Hol	
1	5/17	Mon	543	242	392.1	
2	5/18	Tue	348	203	275.7	
3	5/19	Wed	371	186	278.4	
4	5/20	Thu	469	188	328.4	
5	5/21	Fri	235	105	170.0	
6	5/24	Mon	443	221	332.0	
7	5/25	Tue	360	189	274.6	
8	5/26	Wed	428	224	325.7	
9	5/27	Thu	334	168	251.3	
10	5/28	Fri	214	143	178.6	
11	5/31	Mon	*	*	*	

Time Series Plot

You start your analysis by creating a time series plot for the average overall hold time.

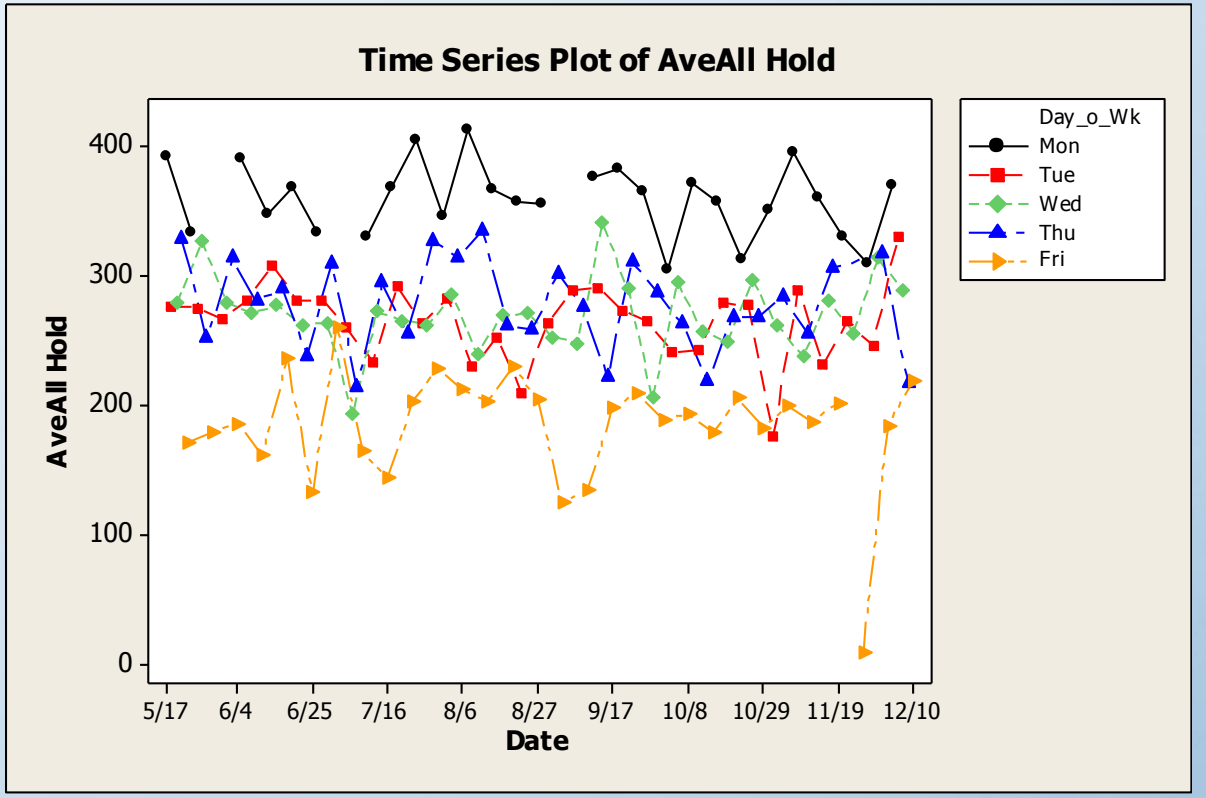
A time series plot is a display of the data points in the order in which they were collected.

Does anything jump out at you?



Stratified Time Series Plot

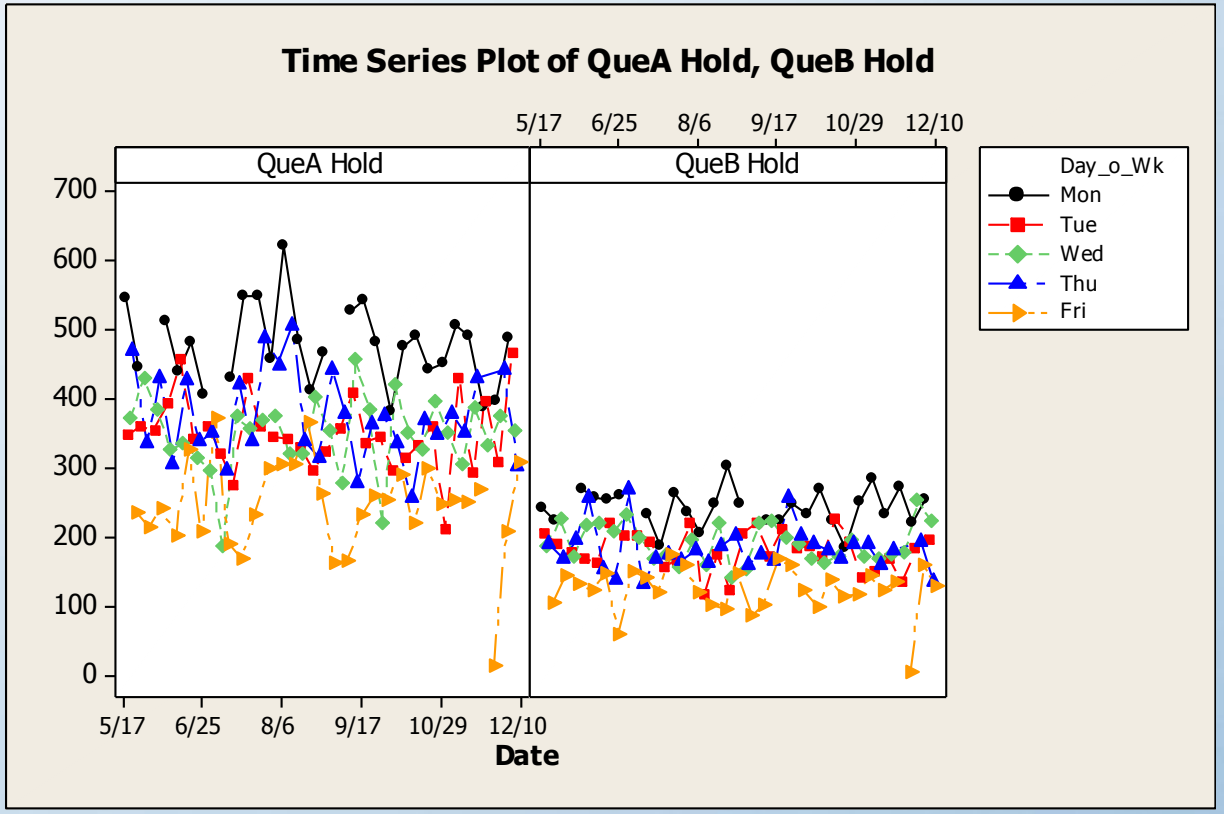
- "Stratifying" means breaking our data into rational groups. Stratification sometimes helps us see clues in the data that might otherwise be obscured.
- You decided to stratify the hold time data by day of the week.
- What do you see here?



Data Comparisons

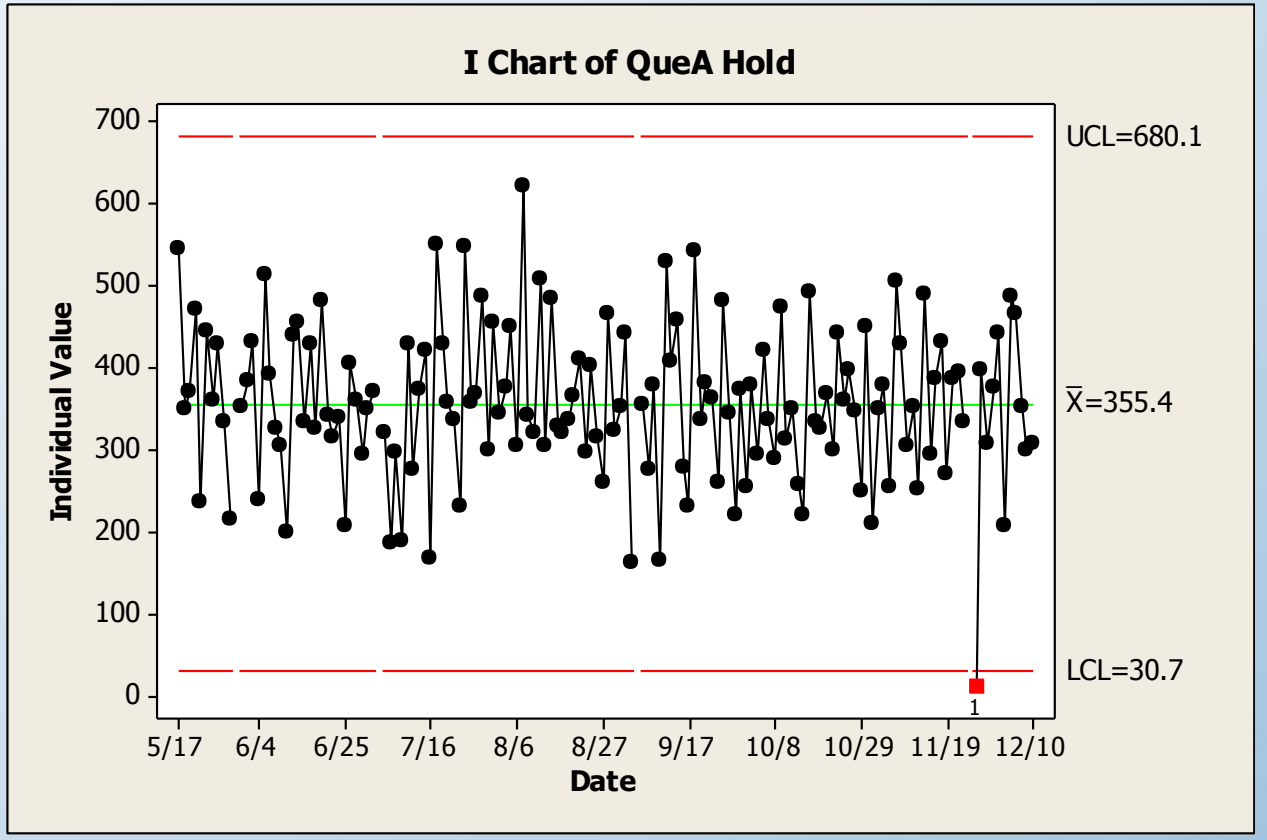
- You can stratify time series plots for Que A and Que B and compare them side by side (making sure the scales are the same).
- What do you know now that you didn't know looking at the overall data?
- What DON'T you know?

The lesson here – plots may not provide all the right answers, but they may lead you to all the right questions!



Control Charts

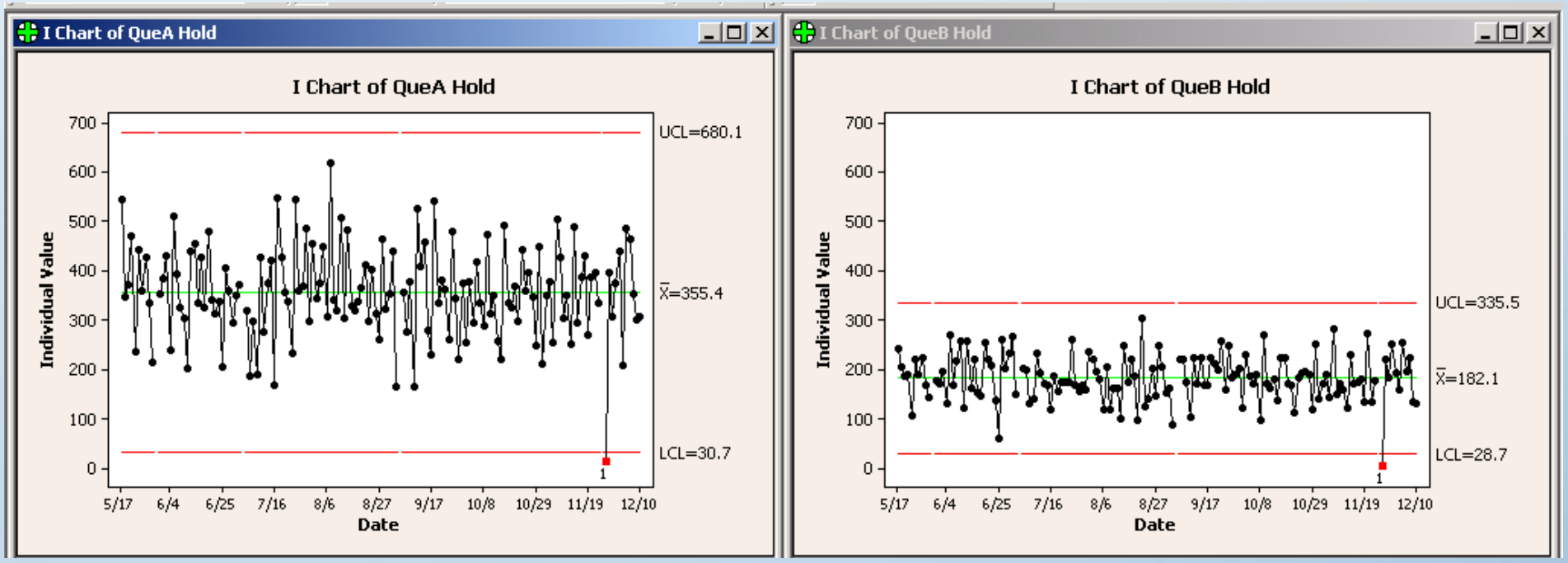
- Are time-ordered plots of results (just like Time Series Plots).
- Use statistically determined control limits that are drawn on the plot.
- Their centerline calculation uses the mean of the data.
- Help you identify special cause versus common cause variation



Control Chart Comparisons

We can compare control charts for Que A and Que B side by side.

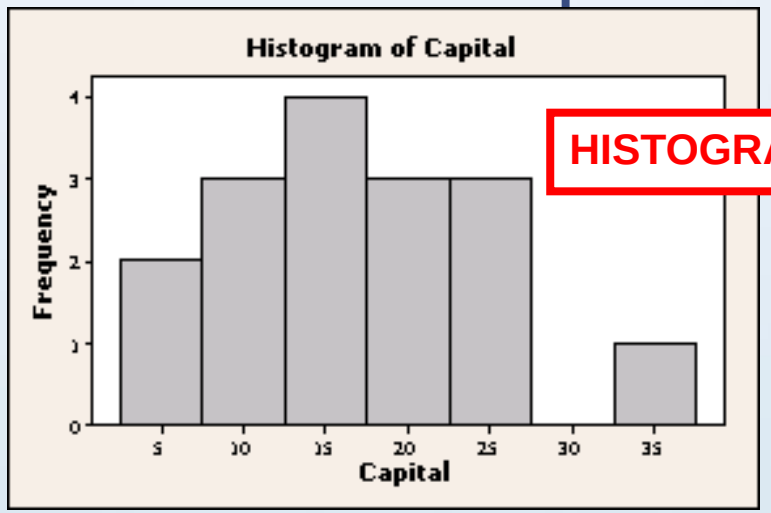
What do you notice?



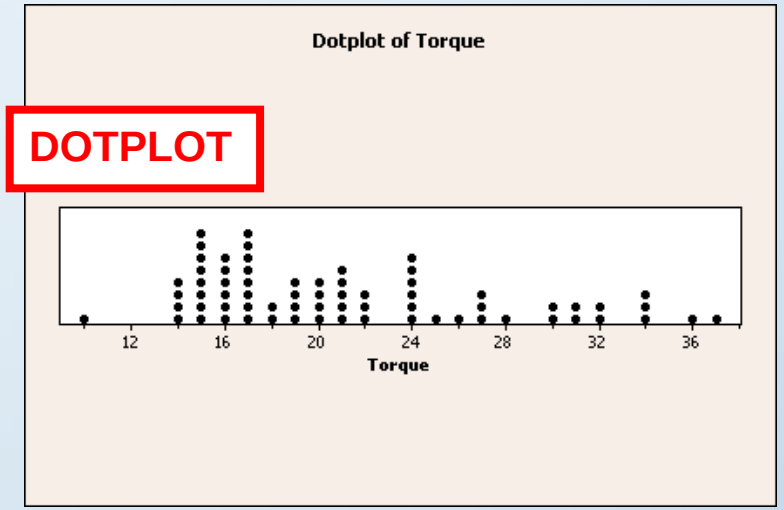
Frequency Plots

- Compare the frequency at which various data values occur
- Reveal the centering, spread, and variation of the data (tabular data does not produce the same insight as a picture)
- Provide a rough indication whether the current process is capable of meeting customer requirements (overlay the specifications on the plot)
- Indicate the underlying statistical distribution of the data (hard to detect patterns with < 50 data points)

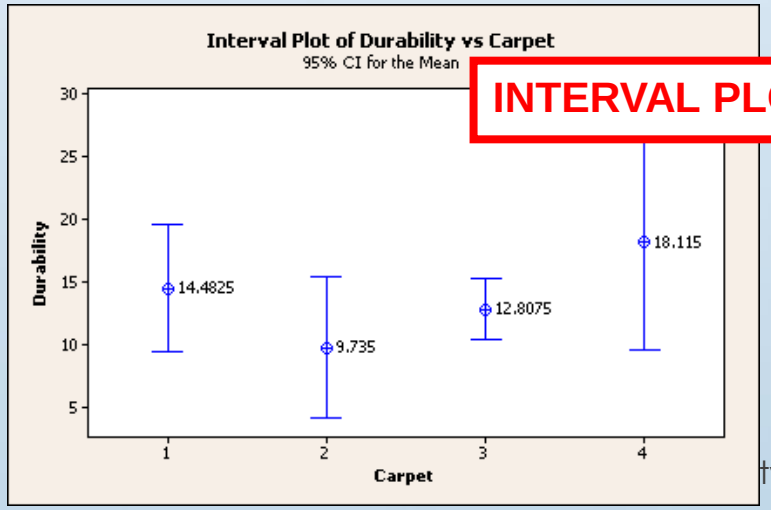
Types of Frequency Plots



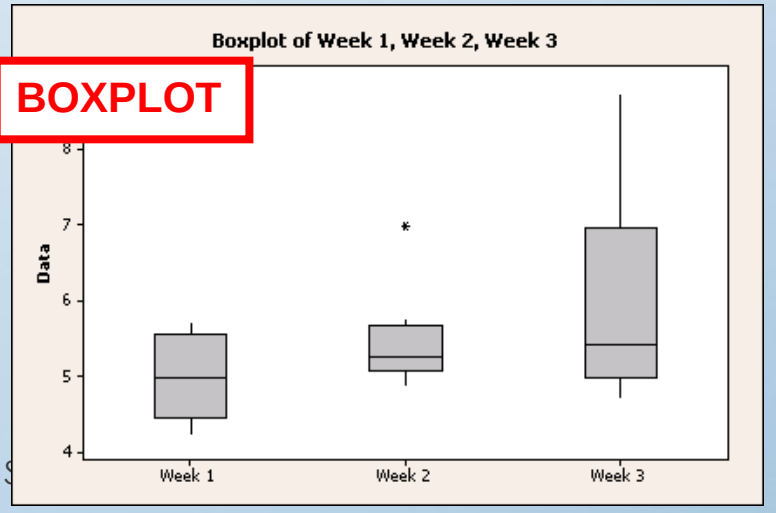
HISTOGRAM



DOTPLOT



INTERVAL PLOT

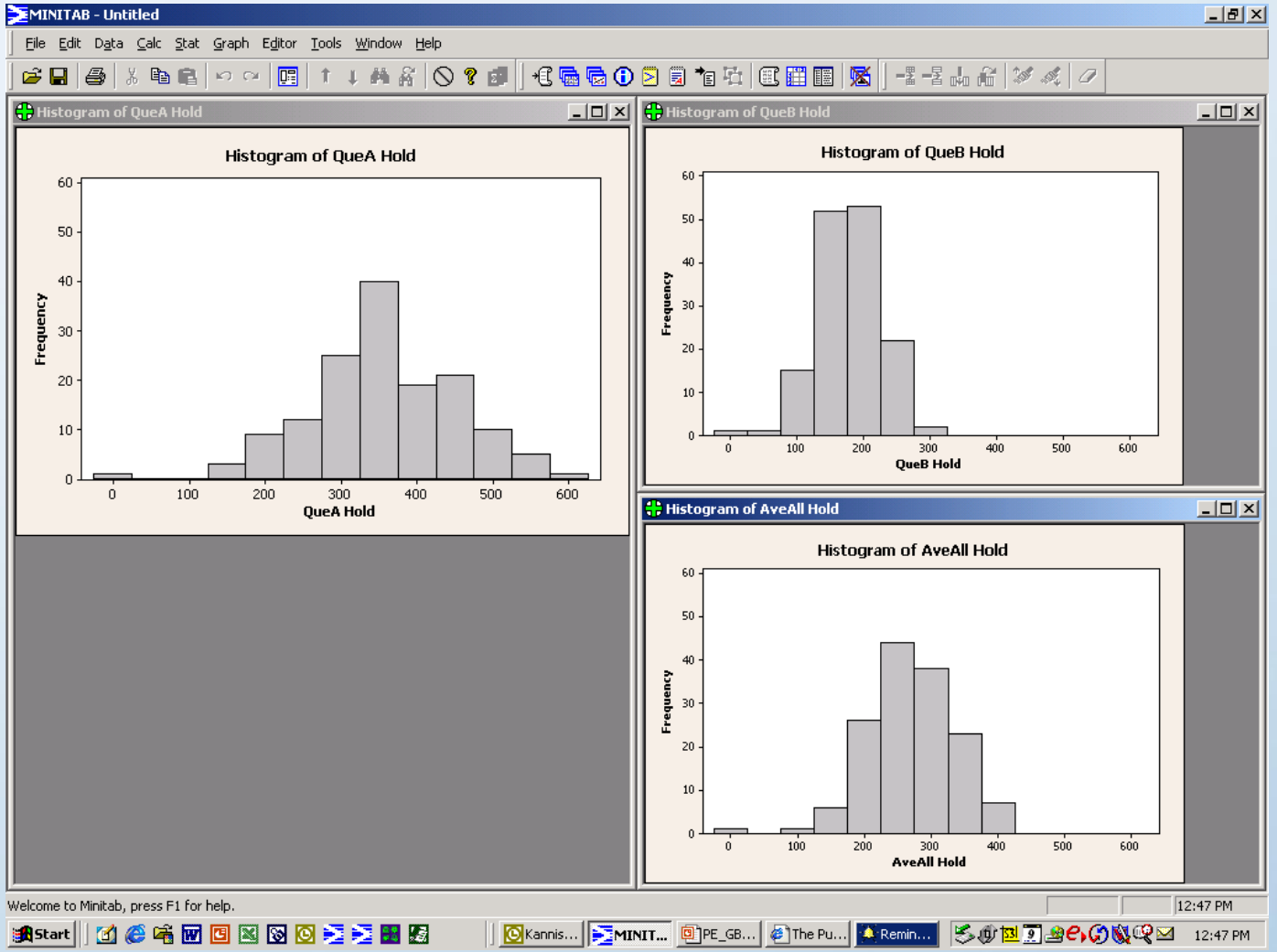


BOXPLOT

Case Study: Histograms

Shape, center, spread – all are clues!

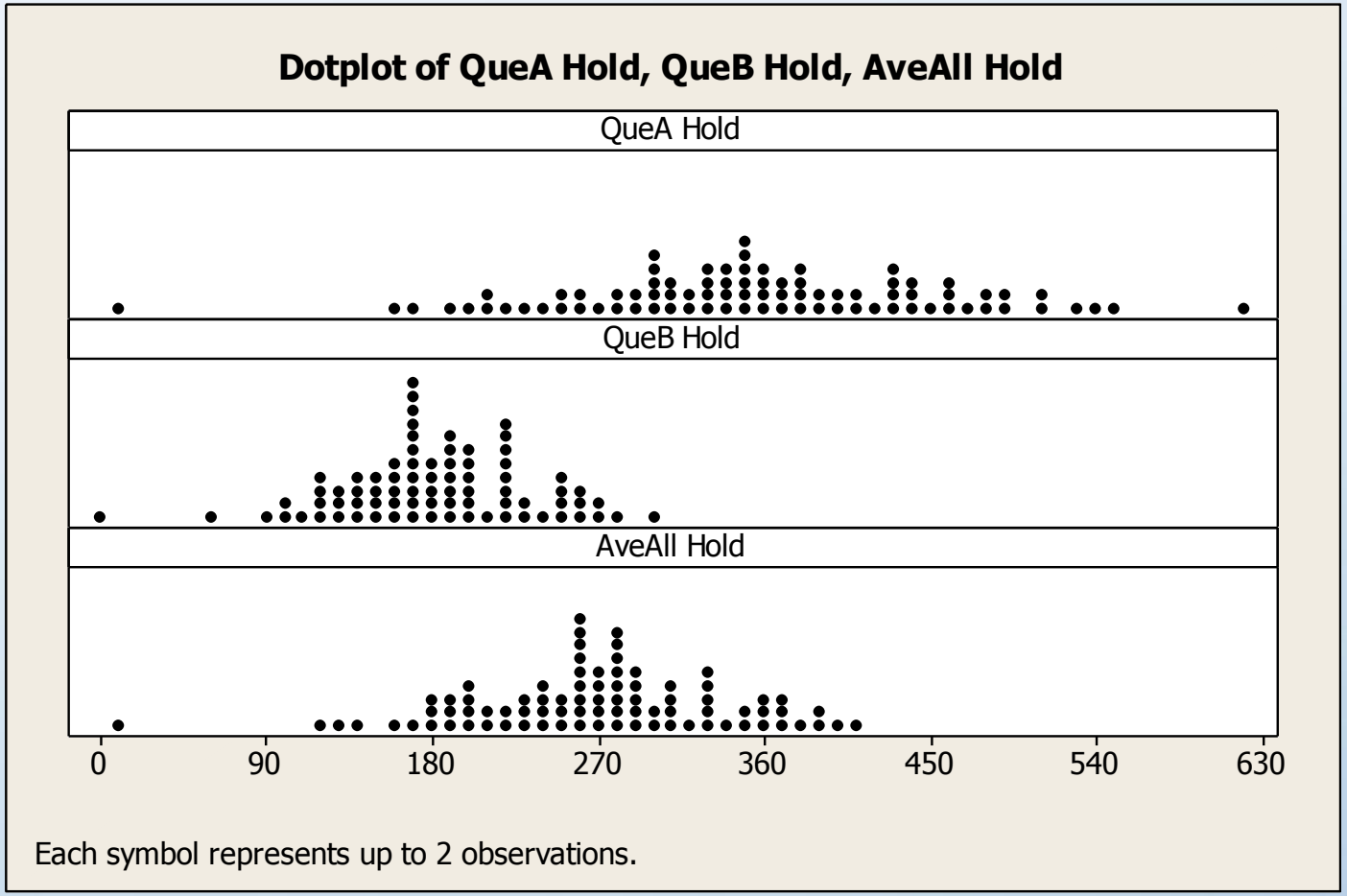
- Shape – is the data roughly normal?
- Center – where is the mean/ median?
- Spread – what is the variability?



Case Study: Dot Plots

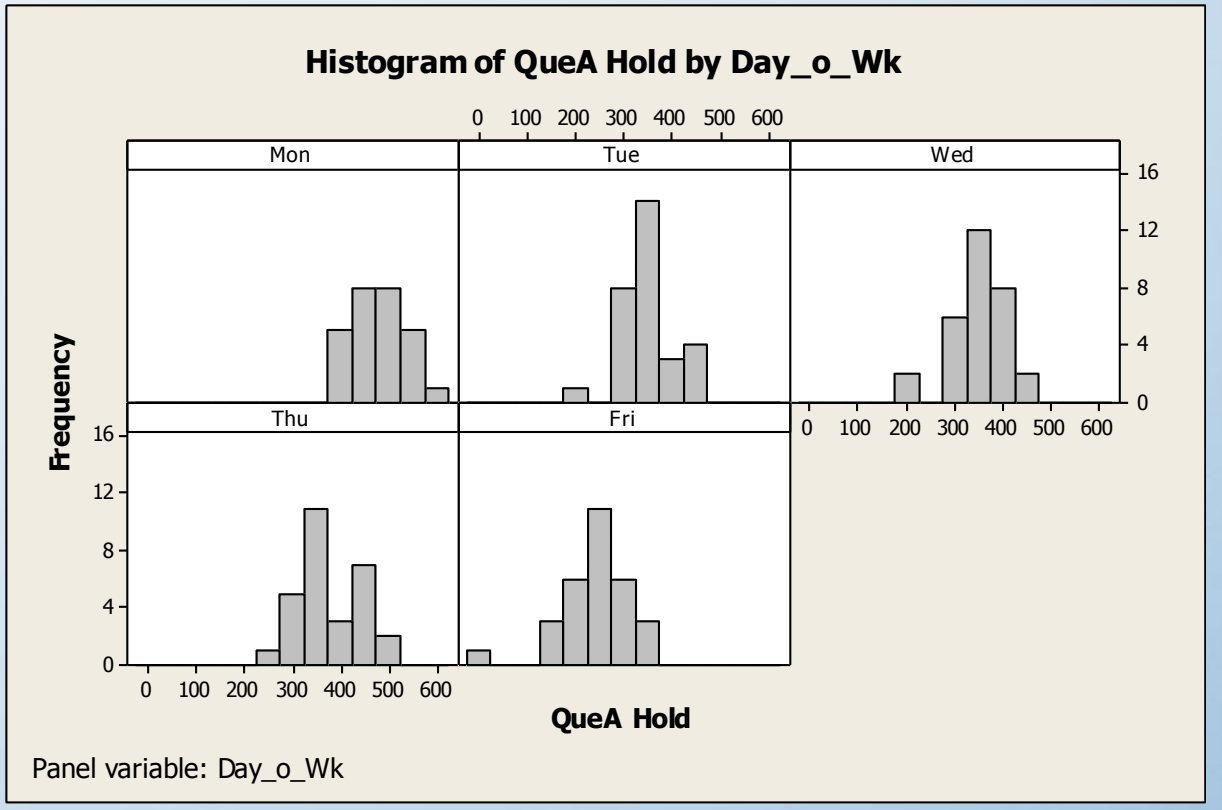
Shape, center, spread again!

A little more granular than a histogram.



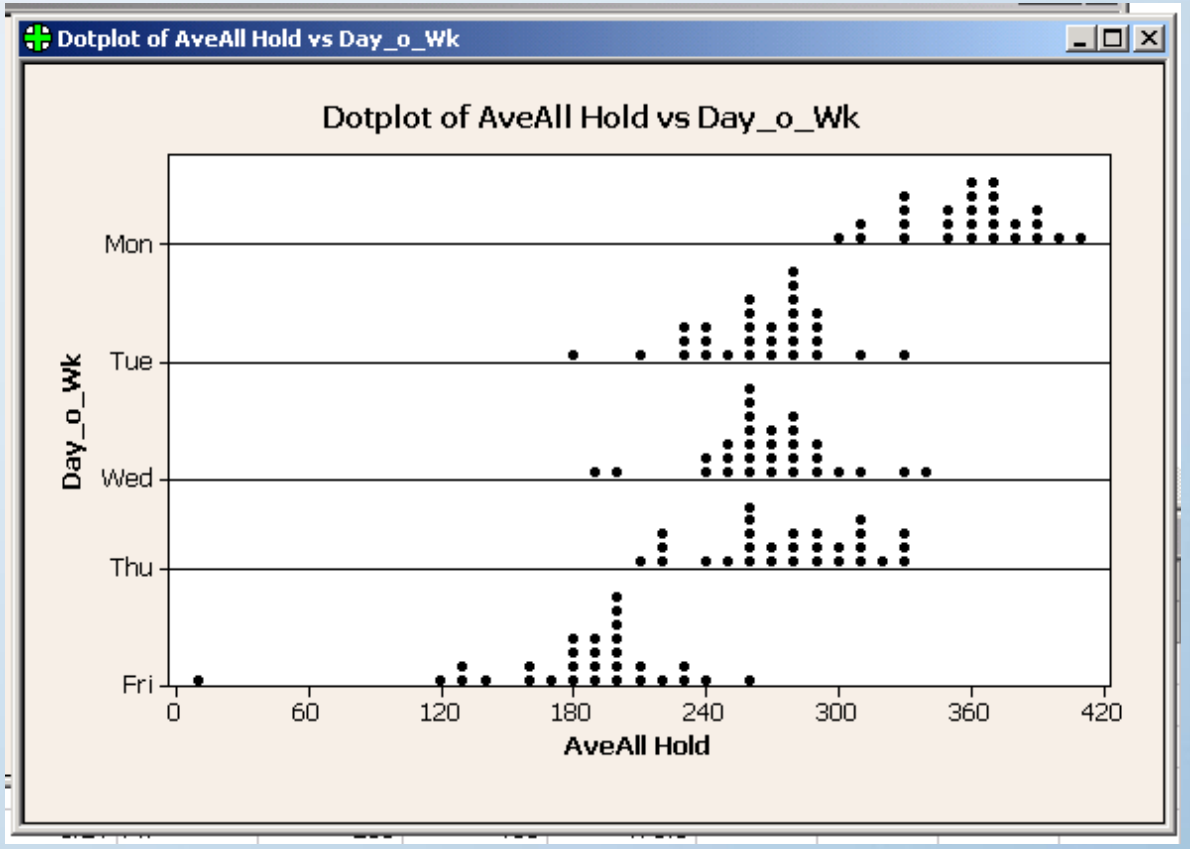
Case Study: Stratified Frequency Plots

- You stratified our Que A hold time by day of week and compared the plots side by side.
- What new clue do you see?
- What new questions do you have?



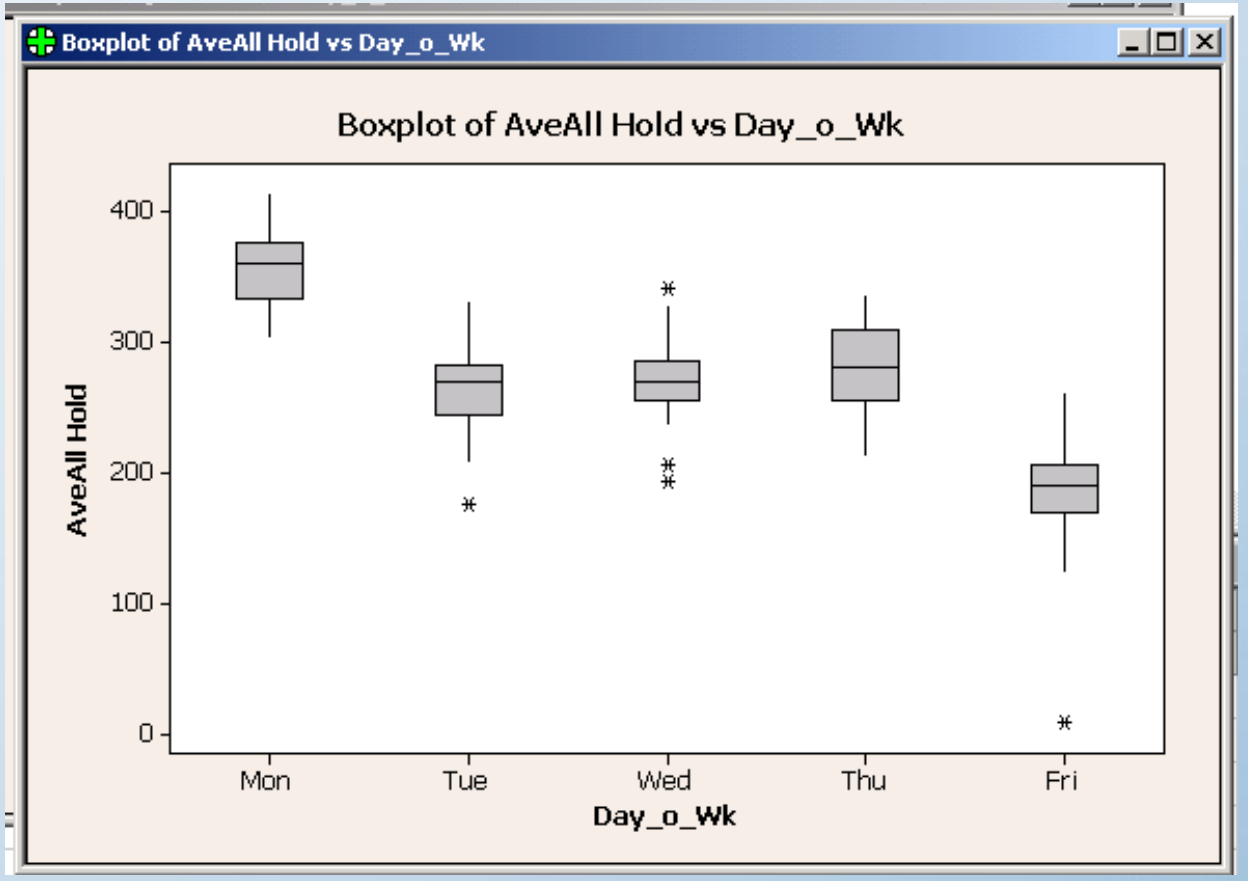
Case Study: Another View

- Let's say we decide to look at a stratified dot plot rather than a stratified histogram.
- There's no one "right" plot to use when we're being data detectives. The "right" plot is the one that gives us new clues or new answers!



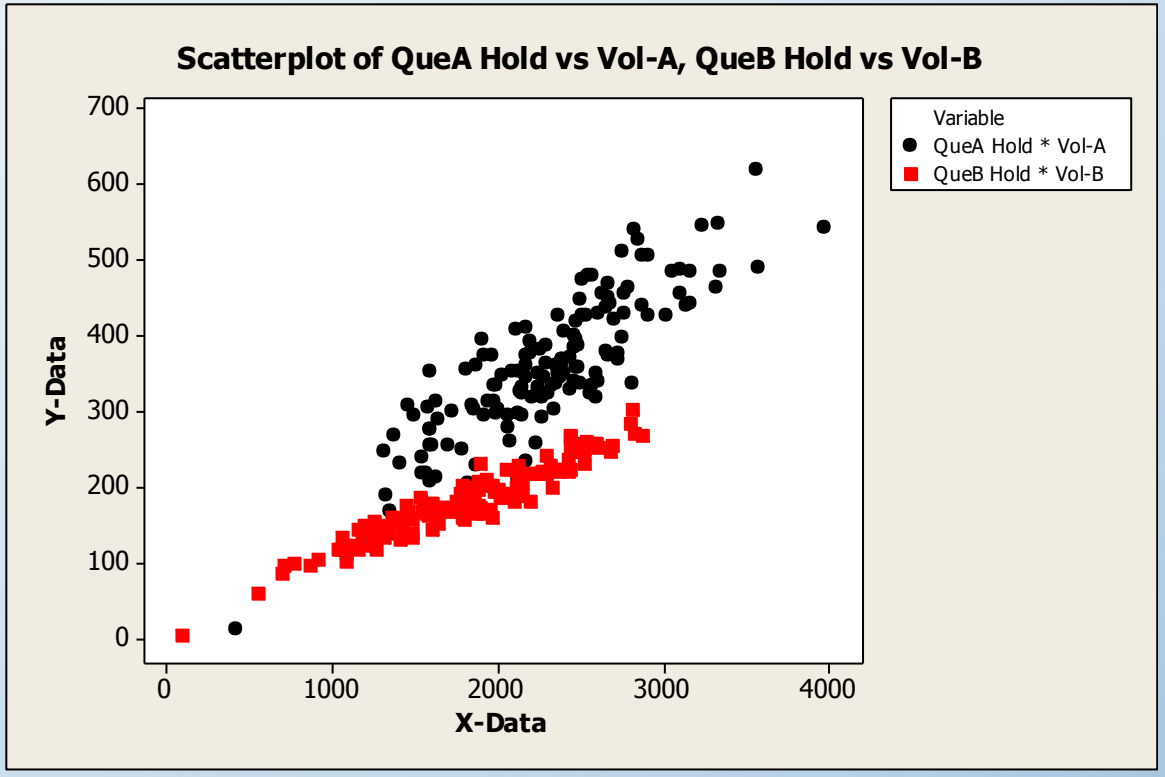
Case Study: Yet Another View

- A stratified box plot provides yet another angle from which we can look at our data
- Again, you can decide which view of the data is the most meaningful to you and your audience!



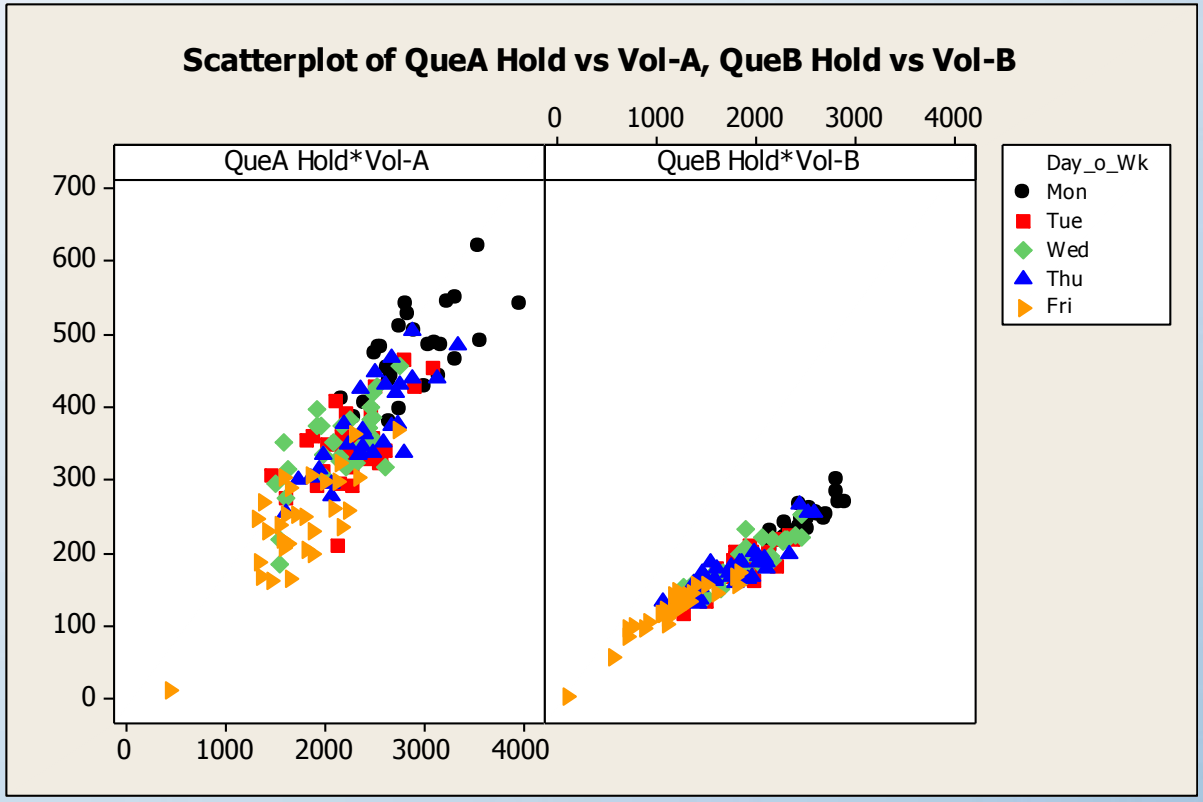
Case Study: Scatter Plots

- Scatter plots enable us to compare relationships between two variables, in this case call volume and hold time.
- Does hold time increase as volume increases?
- What other clues do you see in this stratified scatter plot?



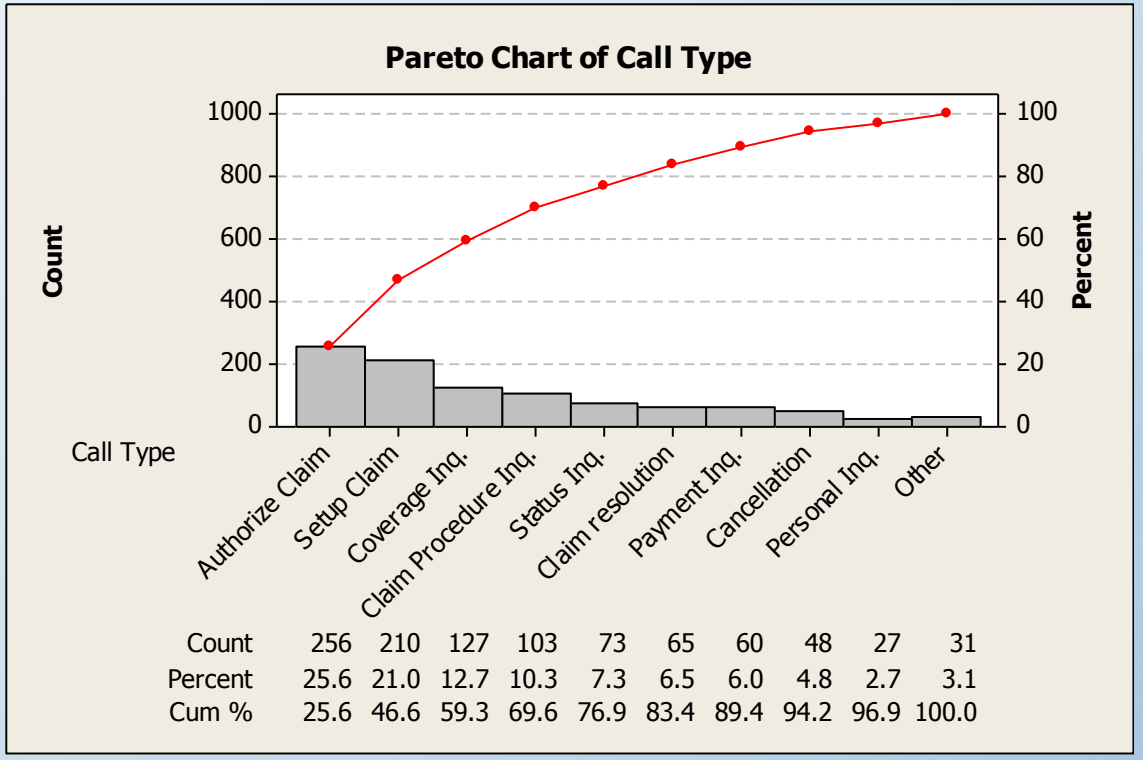
Scatter Plots: Another View

- By "slicing and dicing" our data in different ways, we learn new things about our processes and our problems.
- What does this view help us to see?
- What new questions do we want to ask?



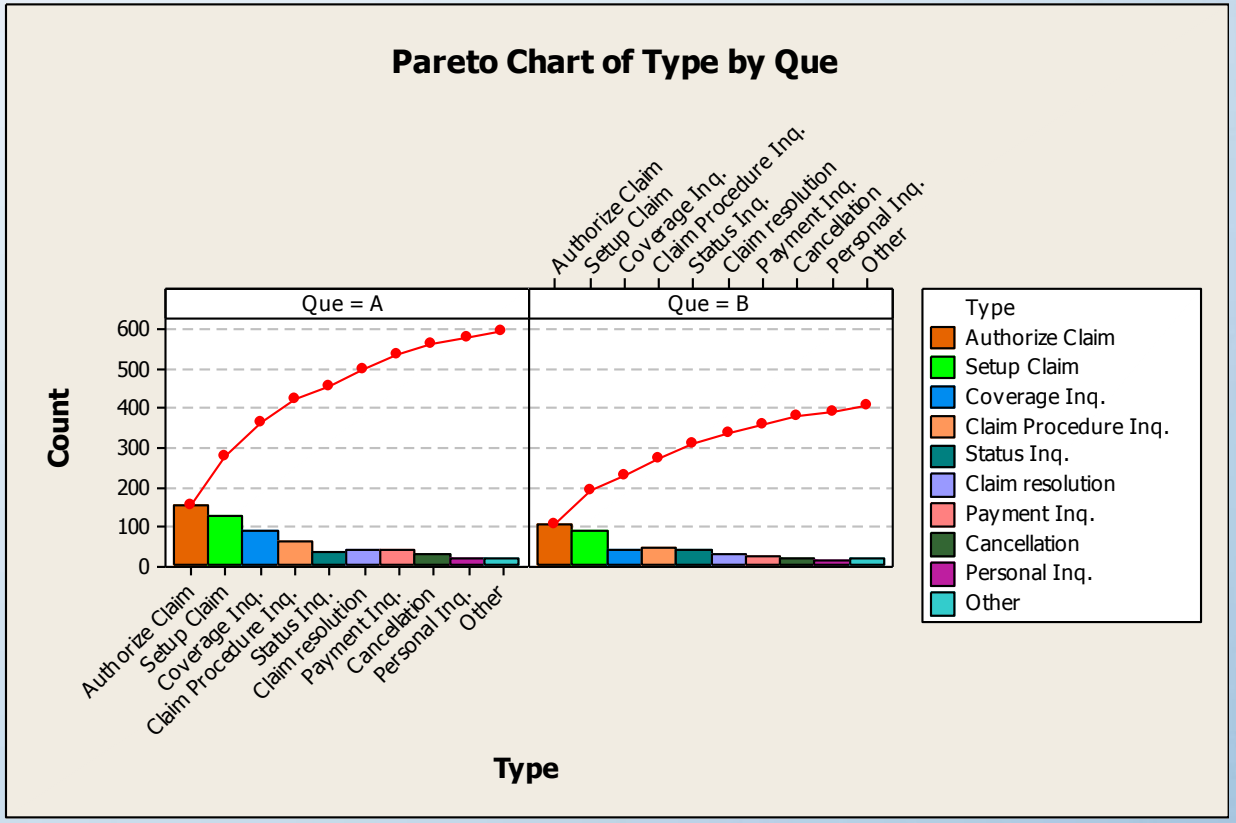
Case Study: Pareto Charts

- A Pareto chart is a graphical tool that helps you break a big problem down into its parts and identify which parts are most important
- It's based on the 80/20 rule – the concept is that 80% of your problem is caused by 20% of the contributors.
- Focus on those key contributors and you can significantly reduce your problem.



Stratified Pareto Charts

- When we stratify our data and look at Que A and Que B side by side, another clue emerges.
- What do you see?



Case Study: Before and After

What We Knew **Before** Looking at Our Data:

- We had a problem: Hold times were too long

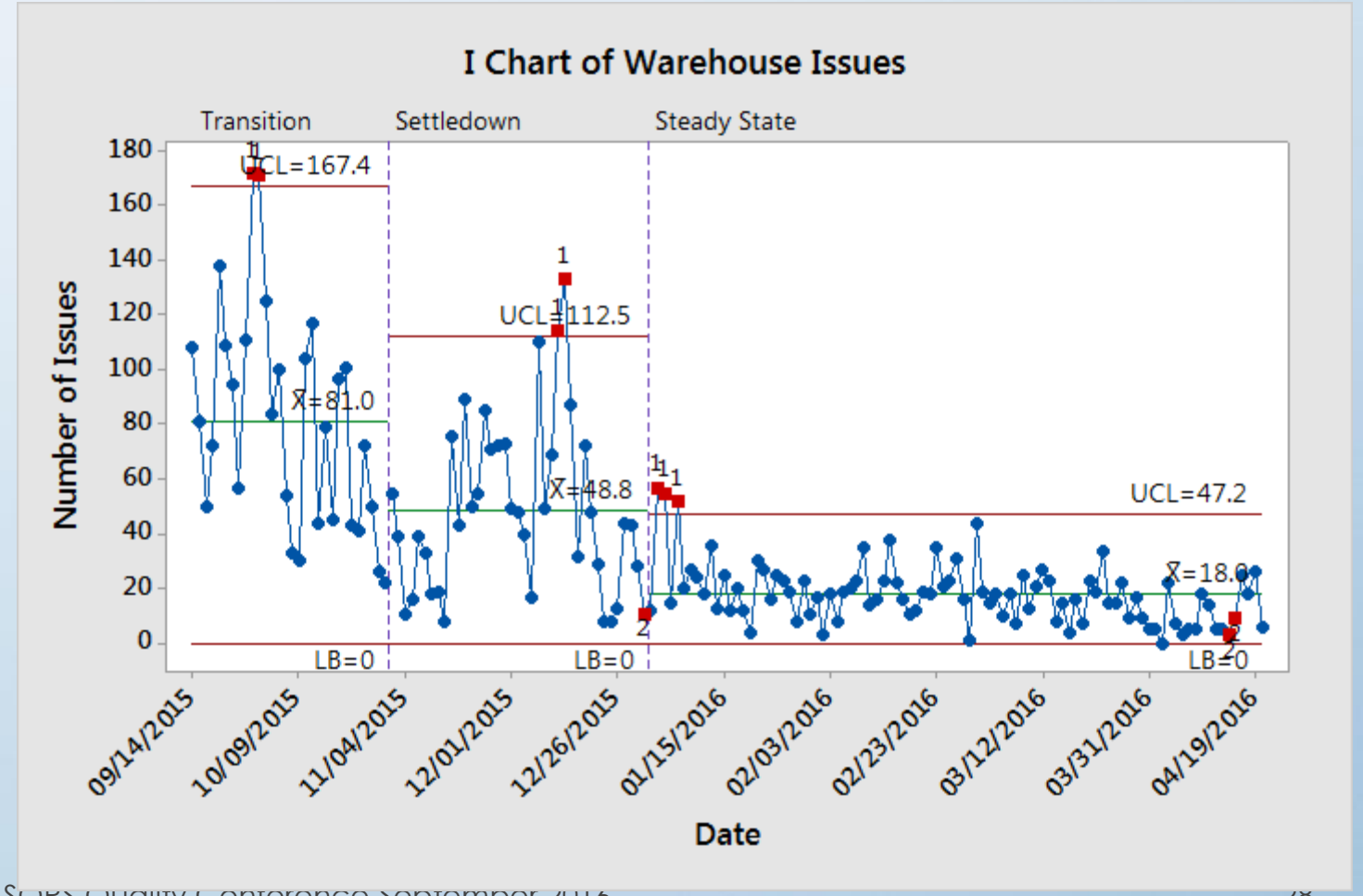
What We Know **After** Looking at Our Data:

- Hold times range from around 60 seconds to 630 seconds
- Hold times are longer on Mondays, shorter on Fridays
- Que A's hold times are consistently longer and more variable than Que B's
- Increasing volume has less effect on Que B than on Que A
- Que A handles more "Authorize Claim" types of calls

Telling a Story Through Data

A Real Life Example from MY Own Work Life:

Data can help you tell a story, demonstrate the impact of improvements, and make a case for change!



The First Three Rules of Data Analysis

- Plot the data
- Plot the data
- Plot the data

Be a data detective!

Use the data to help you find clues that will help you solve your problem!

Questions?

Resources

- Minitab (www.minitab.com)
- **The Visual Display of Quantitative Information** 2nd Edition by Edward R. Tufte
- **Effective Data Visualization: The Right Chart for the Right Data** by Stephanie D. H. Evergreen