

## **ED PIP: Diagnostic Phase**

**Guides: Gaining Insights Through Data**

# Gaining Insights Through Data – Overview

<b>Outcome</b>	<ul style="list-style-type: none"> <li>■ Enhanced ability to learn from the variation that exists in data (Note: one of the key strategies in improvement work is to control or reduce variation)</li> </ul>
----------------	--

<b>Definition: 'What is it?'</b>	<ul style="list-style-type: none"> <li>■ The interpretations that can be made from data vary depending on the chart selected; however, regardless of the tool used, displaying data graphically is imperative for gaining insight from data</li> </ul>
----------------------------------	--

<b>Objectives: 'What is it used for?'</b>	<ul style="list-style-type: none"> <li>■ Four basic tools used to visually display data are described             <ul style="list-style-type: none"> <li>– Run Chart – a simple method for observing variation over time</li> <li>– Frequency Plot – assists in understanding the location, spread, shape and patterns of data</li> <li>– Pareto Chart – helps to identify areas with the greatest opportunity for improvement</li> <li>– Scatter Plot – used to analyse the relationship between two variables</li> </ul> </li> </ul>
---	--

<b>Benefits:</b>	<ul style="list-style-type: none"> <li>■ Depending on your intent, a different graphical tool should be selected; however, because improvement is temporal, a run chart is the most common starting point for learning from data in order to facilitate quality improvement</li> </ul>
------------------	--

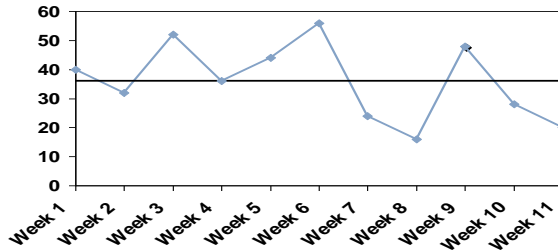
<b>When to use</b>	<ul style="list-style-type: none"> <li>■ As soon as an improvement aim is established, appropriate measures should quickly be identified and data plotted on a run chart</li> </ul>
--------------------	---

***Tip for integrating Lean principles into healthcare:***

- *Part of successfully implementing Lean in healthcare is adopting common language that may have originated in manufacturing and internalizing how it is used in a healthcare environment*
- *There are many ways to present data graphically. Statistical analysis is highly complex, however, there are some basic techniques that can be used to help interpret data. This guide focuses on a handful of these*

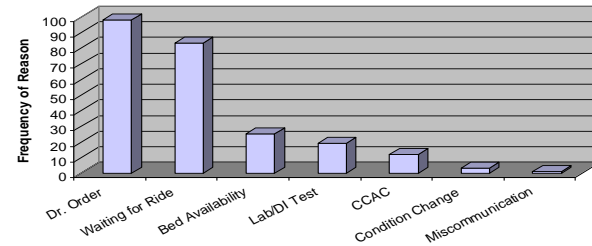
# Gaining Insights through Data – Chart Selection

Run Chart



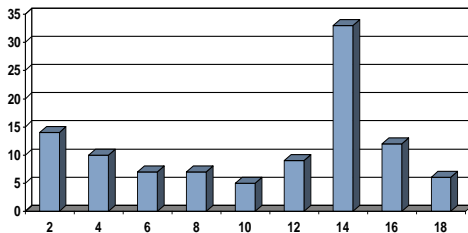
Displays data over time and highlights variation

Pareto Chart



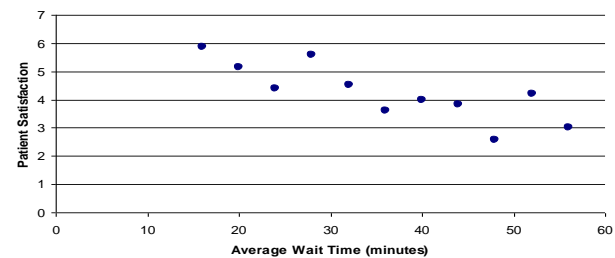
Helps to identify areas with the greatest opportunity for improvement

Frequency Plot



Displays the distribution of observations

Scatter Plot



Makes the relationship between two variables visible

# Three Simple Steps for Constructing a Run Chart

## Step 1: Develop x-axis scale

Develop the horizontal scale for the run chart. This will usually be a time scale, but other alternatives can be used. Appropriate time increments to develop the axis will typically be days, weeks, months, sequential patients, sequential procedures, etc. The scale should cover the time period of interest for the graph, not just the time for which the data are currently available.

## Step 2: Develop y-axis scale

Develop the vertical scale for the run chart. A good scale is one that is easy to plot, easy to read, and leaves ample room for future data that might be larger or smaller than the values used to create the initial run chart. Be sure to construct the vertical scale so that it is high or low enough to encompass variation in future data and reference values such as your goal, a benchmark, or zero if it is meaningful to the chart.

## Step 3: Add a median line

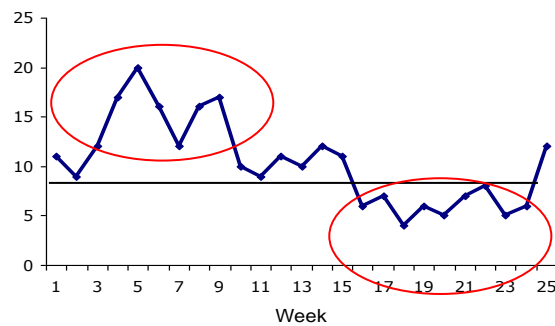
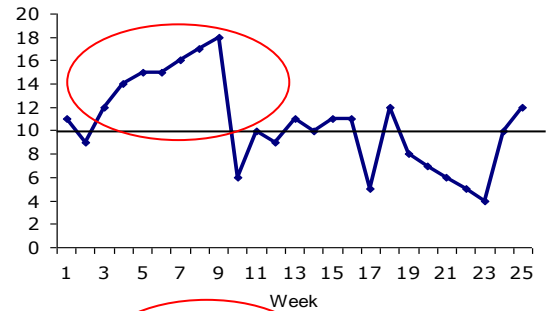
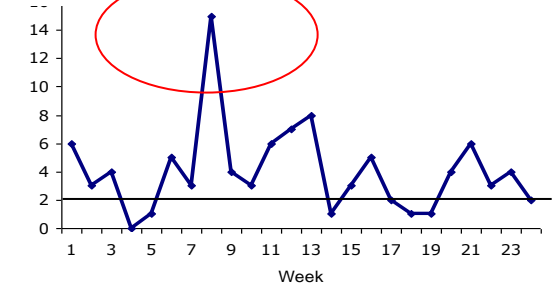
Calculate and place the median of the data on the run chart. The median is the number in the middle of the data set when the data are reordered from the highest to the lowest. Placing the median on a run chart with a small number of data points, typically less than ten, or on a run chart with more than one series of data is not recommended. Finally, add a target line if appropriate. Annotate the chart with unusual events, changes tested etc.

Source: Provost, L. & Murray, S. (2007). "The Data Guide: Learning from Data to Improve Health Care."

Associates in Process Improvement and Corporate Transformation Concepts, pp. 3-3.

# Rules for Identifying Non-Random Signals of Change

Rule
<p><b>1. Shift</b></p> <ul style="list-style-type: none"> <li>• Six or more consecutive data points either all above or all below the median</li> <li>• Note: Skip values that fall on the median and continue counting</li> </ul> <p><b>• Trend</b></p> <ul style="list-style-type: none"> <li>• Five or more consecutive data points all going up or all going down</li> <li>• Note: If the value of two or more consecutive points is the same, ignore one of the points and continue counting</li> </ul> <p><b>3. Astronomical Point</b></p> <ul style="list-style-type: none"> <li>• An astronomical data point is one that is:                     <ul style="list-style-type: none"> <li>➤ An obviously different value</li> <li>➤ Anyone studying the chart would agree that it is unusual</li> </ul> </li> </ul>

Example
 <p>A line chart with a y-axis from 0 to 25 and an x-axis labeled 'Week' from 1 to 25. A horizontal median line is drawn at approximately y=8.5. Two red ovals highlight non-random signals: one around weeks 4-10 where data points are mostly above the median, and another around weeks 16-24 where data points are mostly below the median.</p>
 <p>A line chart with a y-axis from 0 to 20 and an x-axis labeled 'Week' from 1 to 25. A horizontal median line is drawn at approximately y=10. A red oval highlights a trend of five consecutive data points (weeks 4-8) that are all increasing from approximately 12 to 18.</p>
 <p>A line chart with a y-axis from 0 to 14 and an x-axis labeled 'Week' from 1 to 25. A horizontal median line is drawn at approximately y=3. A red oval highlights a single astronomical data point at week 8, which reaches a value of approximately 15, significantly higher than the other points.</p>

Source: Provost, L. & Murray, S. (2007). "The Data Guide: Learning from Data to Improve Health Care." Associates in Process Improvement and Corporate Transformation Concepts, pp. 3-15.

# Detecting Variation Using Run Charts

## Non-Random Signals of Change

- The three simple rules or tests described above are based on probability science and can be used to identify non-random signals of change or special cause.
- All rules should be applied to each graph BUT you do not need to have all three signals of non-random change or all three rules positive to determine that a change has occurred.
- When a “shift” or a “trend” is detected on a run chart there is less than a 5% chance that these tests are positive simply by chance. In other words, the measure displayed on the chart is demonstrating non-random variation.
- Special cause variation is neither good nor bad variation; it simply indicates that the process is unstable and unpredictable. If you detect special cause that represents improvement, for example a reduction in the percent LWBS, you should attempt to learn from this situation and emulate the conditions.
- Variation can result in inefficiencies, waste, rework, ineffective care, errors, and injuries in our healthcare system.

*“If I had to reduce my message for management to just a few words, I’d say it all had to do with reducing variation.”*

– W. Edwards Deming

1/ Provost, L. & Murray, S. (2007). “The Data Guide: Learning from Data to Improve Health Care.” *Associates in Process Improvement and Corporate Transformation Concepts*, pp. 3-14.

2/ Source: Lloyd, R. & Scoville, R. “Better Quality Through Better Measurement.” Presented at the 2008 IHI National Conference. Nashville, Tennessee.

# Gaining Insights Through Data – Frequency Plot

## When and How Is It Used?

- The terms frequency plot and histogram are used interchangeably
- A frequency plot is used to depict frequencies associated with quantitative variables or continuous data
- The resultant graph will allow the end user to understand how the data for the variable under examination are spread or distributed
- A frequency plot can also be used to examine historical data for accuracy and/or determine the impact of process changes
- It is most useful to use a frequency plot when there are 50 or more data points available for analysis

1/ Adams, M. Kiemele, M., Pollock, L., & Quan, T. "Lean Six Sigma: A Tools Guide, 2<sup>nd</sup> Edition." Air Academy Associates, LLC, pp. 111-113.

2/ Provost, L. & Murray, S. (2007). "The Data Guide: Learning from Data to Improve Health Care." Associates in Process Improvement and Corporate Transformation Concepts, pp. 3-3.

# Gaining Insights Through Data – Creating a Frequency Plot

## Step 1

Determine the appropriate number of “cells” or classes for the horizontal axis. A simple trick is to count the total number of data points (n), take the square root of n, and round up as necessary. The scale for the measure of interest should ideally be divided into five to 20 cells which encompass the total range of the measure.

## Step 2

Determine the appropriate width (w) for each cell by dividing the range of the data set by the number of classes as was determined in step one. Place the cells, each of width w, on the horizontal axis (x-axis) and then the number of occurrences or frequency of each value on the vertical axis (y-axis).

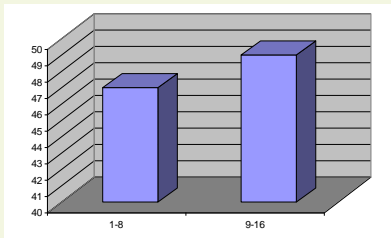
## Step 3

Create a “bar” for each interval according to the number of occurrences of each value. Stratification can be used with a frequency plot by breaking the bar for each interval according to the number of occurrences for each level of the stratification variable.

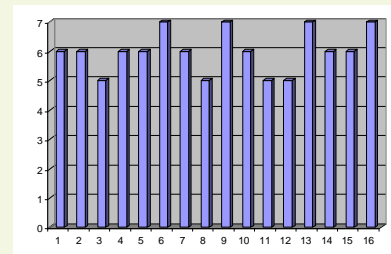


- The appearance of a frequency plot is determined by the number of “cells” selected for the horizontal axis. Too few cells will over summarize the data (that is, provide insufficient detail). On the other hand, too many cells will not adequately summarize the data.

### Too Few Cells:



### Too Many Cells:

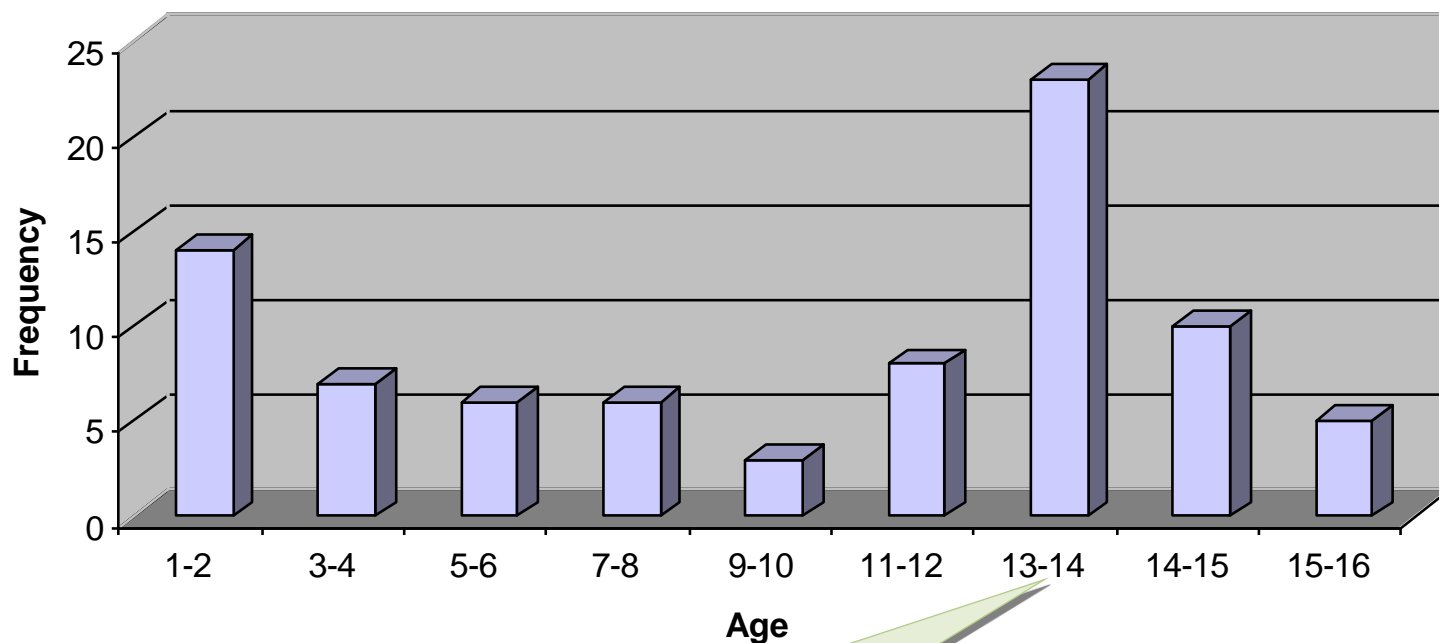


1/ Adams, M. Kiemele, M., Pollock, L., & Quan, T. “Lean Six Sigma: A Tools Guide, 2<sup>nd</sup> Edition.” Air Academy Associates, LLC, pp. 111-113.

2/ Provost, L. & Murray, S. (2007). “The Data Guide: Learning from Data to Improve Health Care.” Associates in Process Improvement and Corporate Transformation Concepts, pp. 4-6.

# Gaining Insights Through Data – Example Frequency Plot

## Histogram: Pediatric Head Injury



- It is apparent that the peak age for head injury among paediatric patients is 13-14 years old
- By learning about the activities of children in this age group, educational or other interventions could be targeted at their specific needs and interests

# Gaining Insights Through Data – Pareto Chart

## When and How Is It Used?

- While a frequency plot is used to depict continuous data, a Pareto chart is used to display the relative frequency of attribute or qualitative data
- Occurrences such as defects, patient complaints, and adverse drug events can often be organized into categories (for example, a Pareto chart could display the contribution of wrong time, wrong dose, omitted medication, wrong patient, wrong medication etc. to the overall occurrences of medication errors on a medical unit)
- Therefore the Pareto chart is key to assisting in the determination of which problem(s) should be addressed first in improvement efforts
- The categories, ordered from most frequent to least, are placed on the horizontal axis and a bar is used to graph the number of occurrences in each category
- The “Pareto effect” refers to situations when 20% of the categories contain 80% of the occurrences



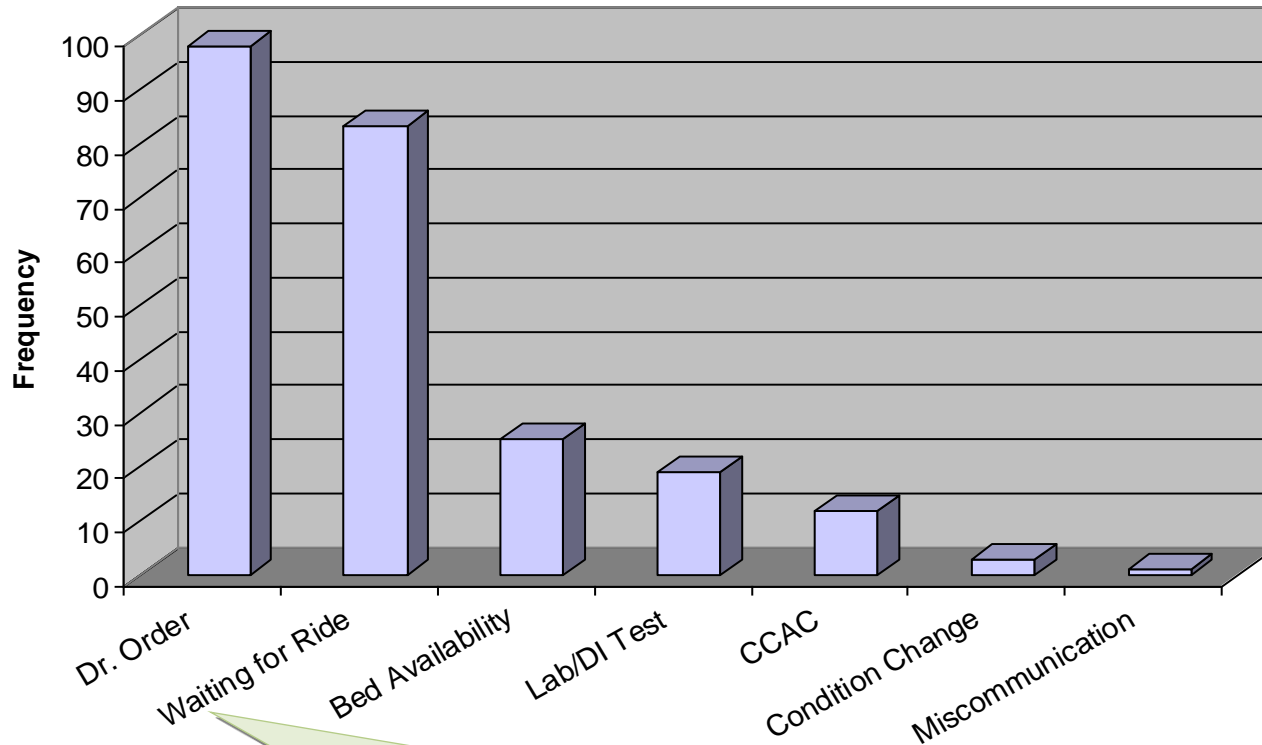
- Data for at least 50 occurrences of the variable of interest should be collected and each occurrence placed into a category
- The most frequently occurring problem is not necessarily the most important. For example, although “wrong time” medication errors may be determined to be the most commonly occurring medication error, the second most frequent error (“medication dose”) might have the most potential for negatively impacting patients

1/ Adams, M. Kiemele, M., Pollock, L., & Quan, T. “Lean Six Sigma: A Tools Guide, 2<sup>nd</sup> Edition.” Air Academy Associates, LLC, pp. 115-117.

2/ Provost, L. & Murray, S. (2007). “The Data Guide: Learning from Data to Improve Health Care.” Associates in Process Improvement and Corporate Transformation Concepts, pp. 4-7 to 4-8.

# Gaining Insights Through Data – Example Pareto Chart

Reason for Discharge Delay



- Discharge is most often delayed because the physician order is not available
- Approximately 75% of discharge delays arise because either the discharge order was not written or the patient does not have transportation available

# Gaining Insights Through Data – Scatter Plot

## When and How Is It Used

- A scatter plot is the graphic display of two variables; more specifically, this graph is used to understand the direction and strength of the relationship between two variables
- The dependent variable is plotted on the y-axis while the independent variable is plotted on the x-axis
- Depending on the pattern depicted on a scatter plot one can determine if there is:
  - A positive relationship between two variables
  - A negative relationship between two variables
  - No correlation between two variables
- One should be mindful that the scaling of the axes is arbitrary and therefore may influence whether a relationship is depicted as being either weak or strong



- Scatter plots do not prove anything, they are simply used to understand relationships
- It is possible to have a high correlation between two variables while no direct cause and effect relationship exists; it is important to remember that multiple variables may be absent from the data which would be required to determine the true cause and effect relationship
- Scatter plots are most useful when there are 50 or more data points available for analysis

1/ Adams, M. Kiemele, M., Pollock, L., & Quan, T. "Lean Six Sigma: A Tools Guide, 2<sup>nd</sup> Edition." Air Academy Associates, LLC, pp. 125-128.

2/ Provost, L. & Murray, S. (2007). "The Data Guide: Learning from Data to Improve Health Care." Associates in Process Improvement and Corporate Transformation Concepts, pp. 4-4 to 4-9.

# Gaining Insights Through Data – Sample Scatter Plots

## Relationship Between Variables

### Positive

- In the sample scatter plot to the right, time spent with healthcare providers is associated with an increase in satisfaction scores

### Negative

- In the sample scatter plot provided, patient satisfaction scores are compromised as wait times increase

### No Correlation

- The dependent variable does not appear to be associated with the independent variable in question

## Example

