

# AQS 200

## ROOT CAUSE INVESTIGATION

This material is based upon work supported  
by the National Science Foundation under  
Grant No. 1304474



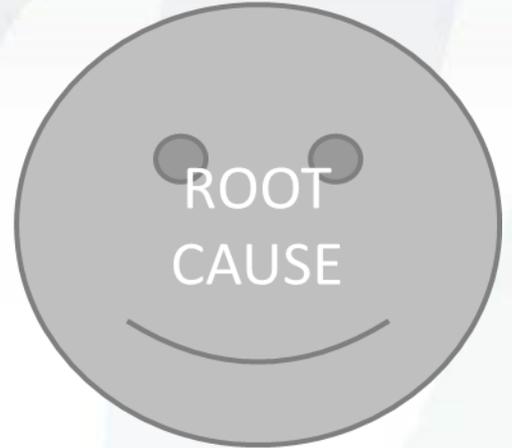
*Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.*

# Lecture 8

## Tools for Data Analysis – Part 1

- Histograms
- Pareto Charts
- Scatter Charts





Problem/Issue Identified

Understand the problem

Investigate potential causes

Collect data on potential causes

Analyze data collected

Identify Root Cause

Eliminate Problem

Implement Solution

Understanding the problem

- Flowchart
- Critical Incident
- Spider Chart
- Performance Matrix

Investigate Potential Causes

- Brainstorming
- Brain-writing
- Is – Is Not Matrix
- Nominal Group Technique
- Paired Comparison

Data Collection

- Sampling
- Surveys
- Check Sheets

# HISTOGRAM

# HISTOGRAM

- Used to display the distribution and variation of a data set
- Pictorial display of how often (frequency) a value occurs within a data set
- Sometimes called bar chart - but they are different

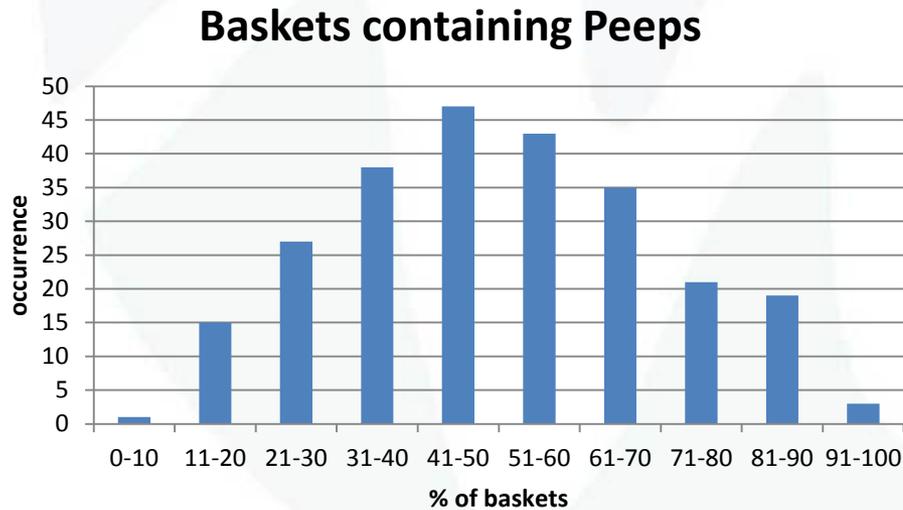
# HISTOGRAM

- Used to display the distribution and variation of a data set
- Pictorial display of how often (frequency) a value occurs within a data set
- Sometimes called bar chart - but they are different
  - Bar chart depicts categories



# HISTOGRAM

- Used to display the distribution and variation of a data set
- Pictorial display of how often (frequency) a value occurs within a data set
- Sometimes called bar chart - but they are different
  - Bar chart depicts categories
  - Histogram depicts frequency of occurrence

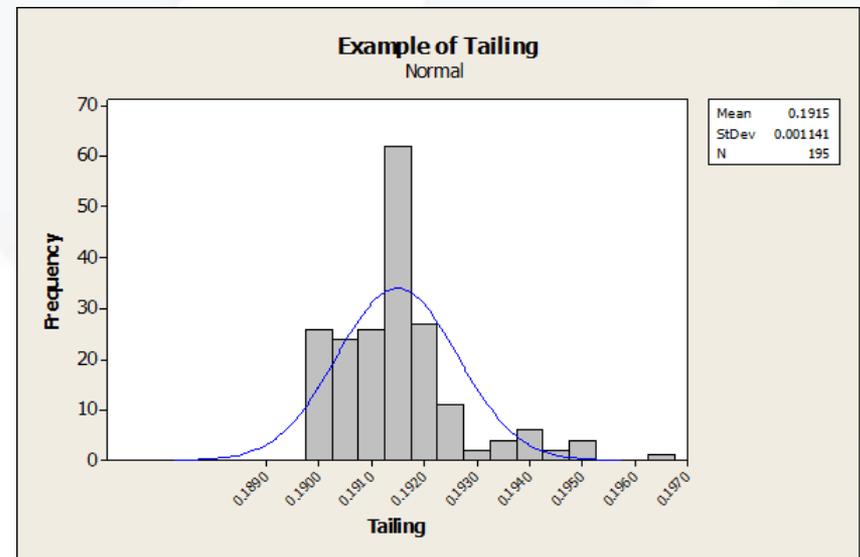
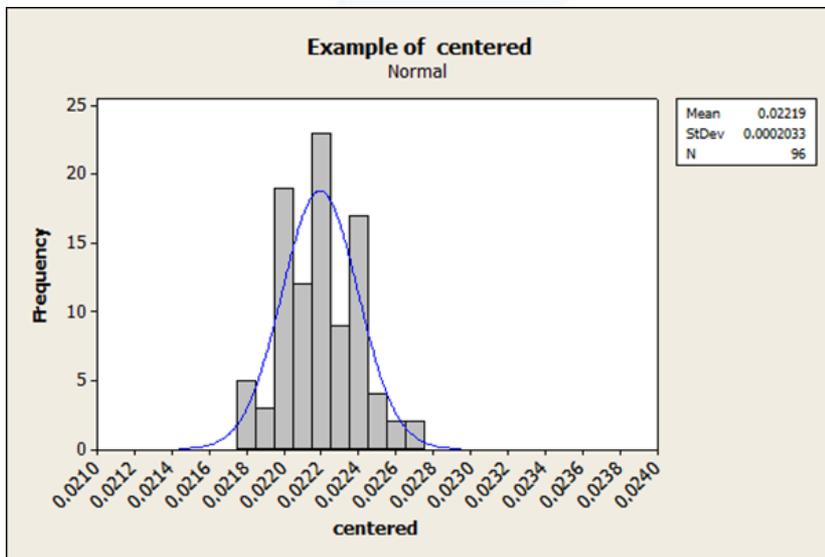


# HISTOGRAM

- Purpose & Applications
  - Presenting data to determine which causes dominate
  - Understanding the distribution of occurrences
- When to use a histogram
  - Data is numerical (not pass/fail)
  - Determine whether process distribution is normal
  - Determine whether process meets customer requirements
  - Measuring supplier process output
  - Are two process outputs the same
- Patterns provide clues

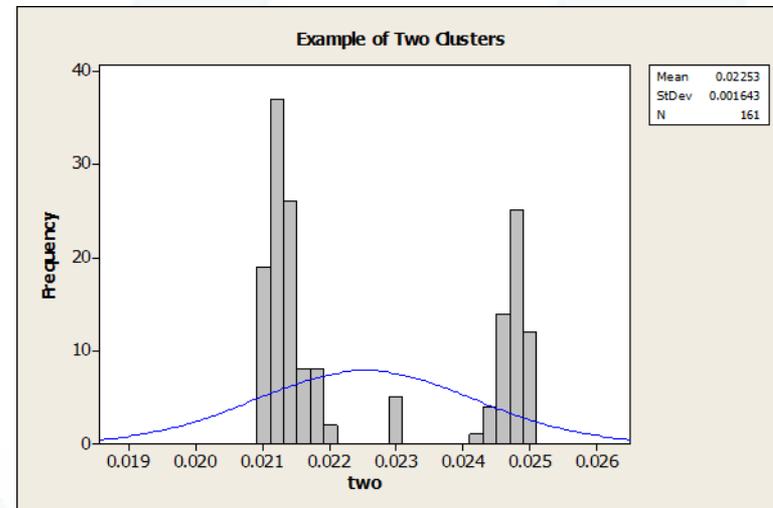
# HISTOGRAM

- Patterns
  - One peak
    - Shows central tendency, demonstrates position of the mean
    - If mean is off-center need to understand why
    - Tailing



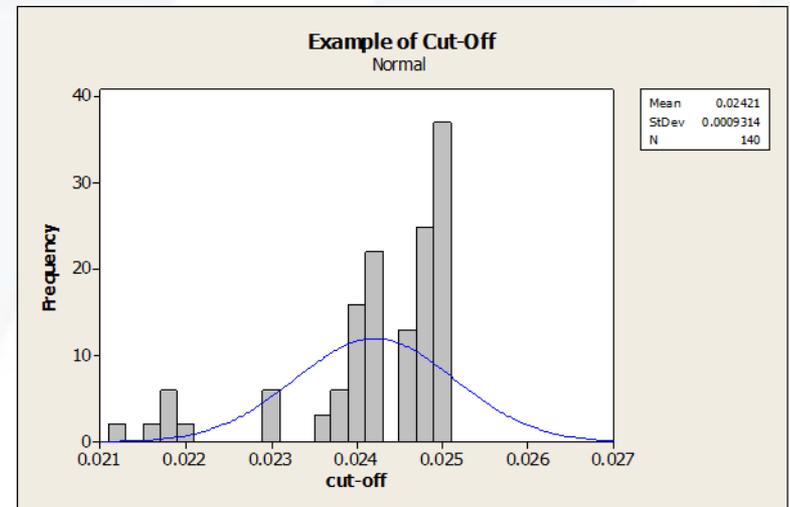
# HISTOGRAM

- Patterns
  - One peak
    - Demonstrates position of the mean
  - Two peak
    - Could be two different data sources
      - Multiple operators
      - Shifts
      - Suppliers



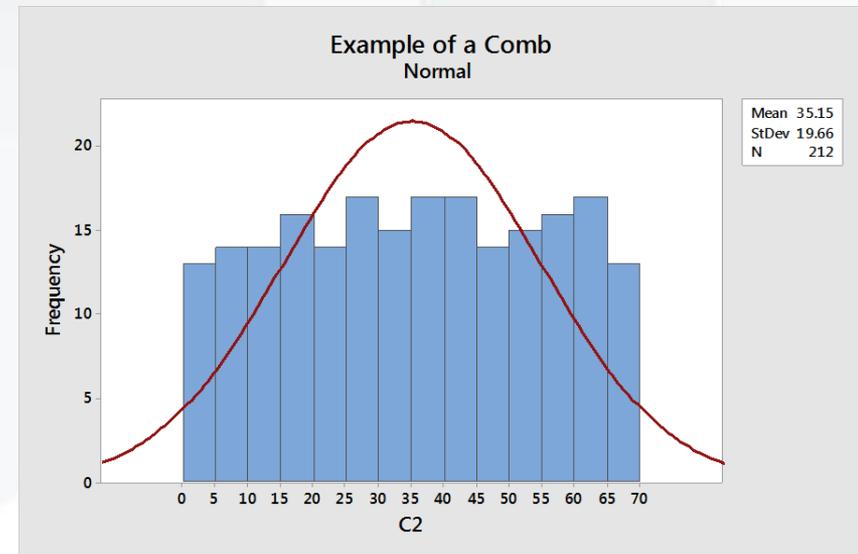
# HISTOGRAM

- Patterns
  - One peak
    - Demonstrates position of the mean
  - Two peak
    - Could be two different data sources
  - **Cut-off**
    - **Without tapering (tailing) suggests selection process**



# HISTOGRAM

- Patterns
  - One peak
    - Demonstrates position of the mean
  - Two peak
    - Could be two different data sources
  - Cut-off
    - Without tapering (tailing) suggests selection process
  - Comb-like Effect
    - Too many classes (bins) defined
      - Data has been divided to where differences cannot be distinguished



# Histograms – How To

1. Count the number of data points  $N$ .

a) at least 30 data points are needed for valid histogram

	# Jellybeans				
1	10	23	94	82	18
2	39	55	69	9	80
3	64	69	15	77	30
4	30	46	94	64	97
5	43	71	16	51	99
6	38	57	55	31	41
7	64	25	29	37	68
8	54	31	50	12	47
9	72	58	85	83	12
10	46	2	47	46	63
11	48	38	10	77	83
12	37	81	98	16	69
13	74	39	76	69	84
14	19	75	85	81	43
15	81	5	95	33	4
16	46	73	17	35	22
17	88	33	63	53	9
18	59	15	49	78	18
19	65	41	98	40	64
20	5	23	35	12	9

# Histograms – How To

1. Count the number of data points  $N$ .
2. Calculate the range  $R$  between the largest and smallest values in the data.

	# Jellybeans				
1	10	23	94	82	18
2	39	55	69	9	80
3	64	69	15	77	30
4	30	46	94	64	97
5	43	71	16	51	99
6	38	57	55	31	41
7	64	25	29	37	68
8	54	31	50	12	47
9	72	58	85	83	12
10	46	2	47	46	63
11	48	38	10	77	83
12	37	81	98	16	69
13	74	39	76	69	84
14	19	75	85	81	43
15	81	5	95	33	4
16	46	73	17	35	22
17	88	33	63	53	9
18	59	15	49	78	18
19	65	41	98	40	64
20	5	23	35	12	9

$N = 100$    Smallest = 2   Largest = 99  
 $R = 97$

# Histograms – How To

## 3. Determine number of bins

- a) Bins = number classes/categories that cover the entire range of values
- b) The number of bins created is based on total number of data points and type of data

The number of data points $N$	The number of classes $C$
Less than 50	5 to 7
50 to 100	6 to 10
100 to 250	7 to 12
More than 250	10 to 20

Note: book uses the word “Classes” I prefer “bins”

# *Histograms – How To*

3. Determine number of bins
  - a) Bins = number classes/categories that cover the entire range of values
4. Determine width (range within the bin) for each bin
$$H = R/C$$
 where  $H = \text{width}$ ,  $R = \text{range}$   $C = \# \text{ bins}$
5. Using the width divide the data range into the selected number of bins.
  - a) The bins will count sequentially, but evenly spread

# Histograms – How To

- Determine number of bins
  - Bins = number classes/categories that cover the entire range of values
- Determine width (range within the bin) for each bin

$$H = R/C$$

where H = width

R = range; C = # bins

- Using the width divide the data range into the selected number of bins.
  - The bins will count sequentially, but evenly spread

	# Jellybeans				
1	10	23	94	82	18
2	39	55	69	9	80
3	64	69	15	77	30
4	30	46	94	64	97
5	43	71	16	51	99
6	38	57	55	31	41
7	64	25	29	37	68
8	54	31	50	12	47
9	72	58	85	83	12
10	46	2	47	46	63
11	48	38	10	77	83
12	37	81	98	16	69
13	74	39	76	69	84
14	19	75	85	81	43
15	81	5	95	33	4
16	46	73	17	35	22
17	88	33	63	53	9
18	59	15	49	78	18
19	65	41	98	40	64
20	5	23	35	12	9

Lowest value = 2, highest = 99

Range = 97 N = 100

6 bins would be  $97/6 = 17$  per bin

2 – 20, 21 – 39, 40 – 58, 59 – 67, 68 – 79, 80 – 99

10 bins would be  $97/10 = 10$  per bin

2-12, 13-23, 24-34 ..... 90-99

# Histograms – How To

6. To simplify the construction of the histogram, insert the data into a check sheet.
7. Construct the histogram based on the check sheet.
  - a) Mark the bins (classes) along the horizontal axis and the frequency along the vertical.
  - b) Use vertical bars to indicate the distribution among classes.

# *Histograms*

**EXERCISE**

# Pareto Charts

- Pareto Principle (aka 80-20 rule)

“... is a prediction that 80% of effects come from 20% of causes”

-- 1906, Vilfredo Pareto an Italian economist

- The 20% are called “*the vital few*”.

# Pareto Charts

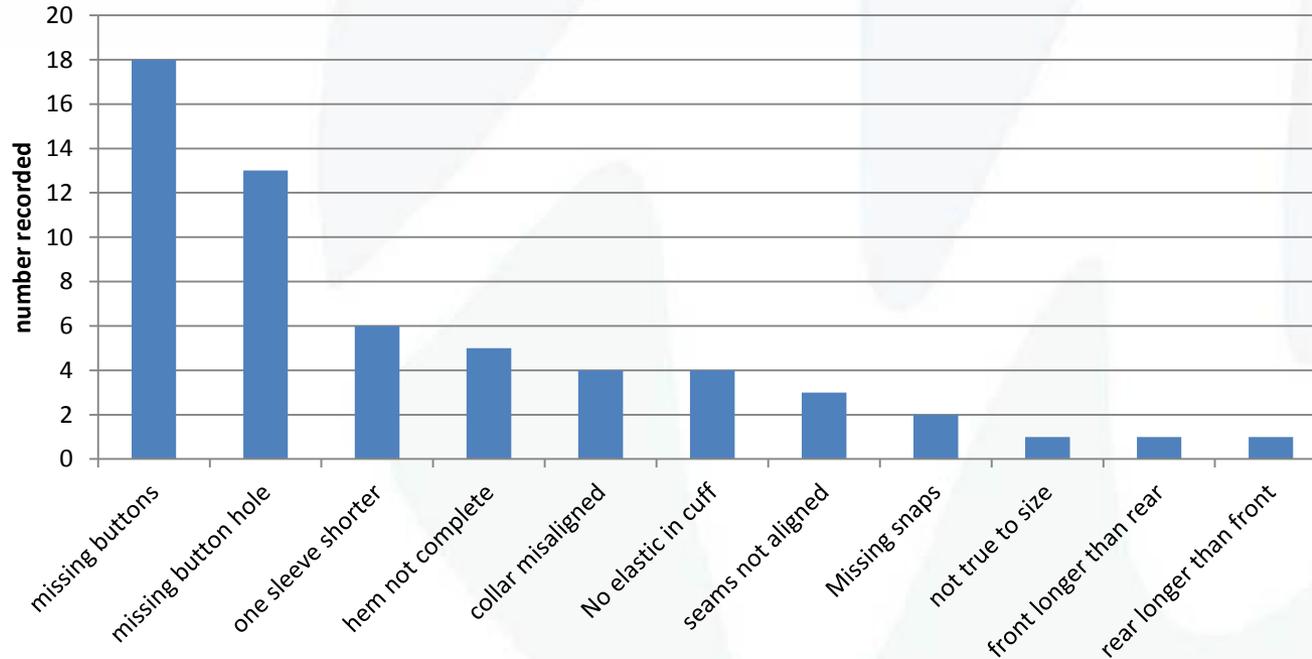
- Graphical display showing causes of a problem sorted by their degree of seriousness
  - A form of a bar chart
  - Similar to histogram, but typically categories rather than dimensions
    - NCR findings
    - Customer Complaints
    - Audit findings (i.e. process, documents, training, etc)
  - Provides visual regarding rank of the problems or causes

**Evaluate the vital few**

# Pareto Charts

- Graphical display showing causes of a problem sorted by their degree of seriousness
  - Similar to histogram, but typically categories rather than dimensions
  - Provides visual regarding rank of the problems or causes

## Q4 - Blouse Defects



Evaluate the vital few

# Pareto Charts

- Graphical display showing causes of a problem sorted by their degree of seriousness
  - Similar to histogram, but typically categories rather than dimensions
  - Provides visual regarding rank of the problems or causes
- **Application**
  - Obtain clearer picture of causes based on importance (occurrence)
  - Determine further investigation

# Pareto Charts – How To

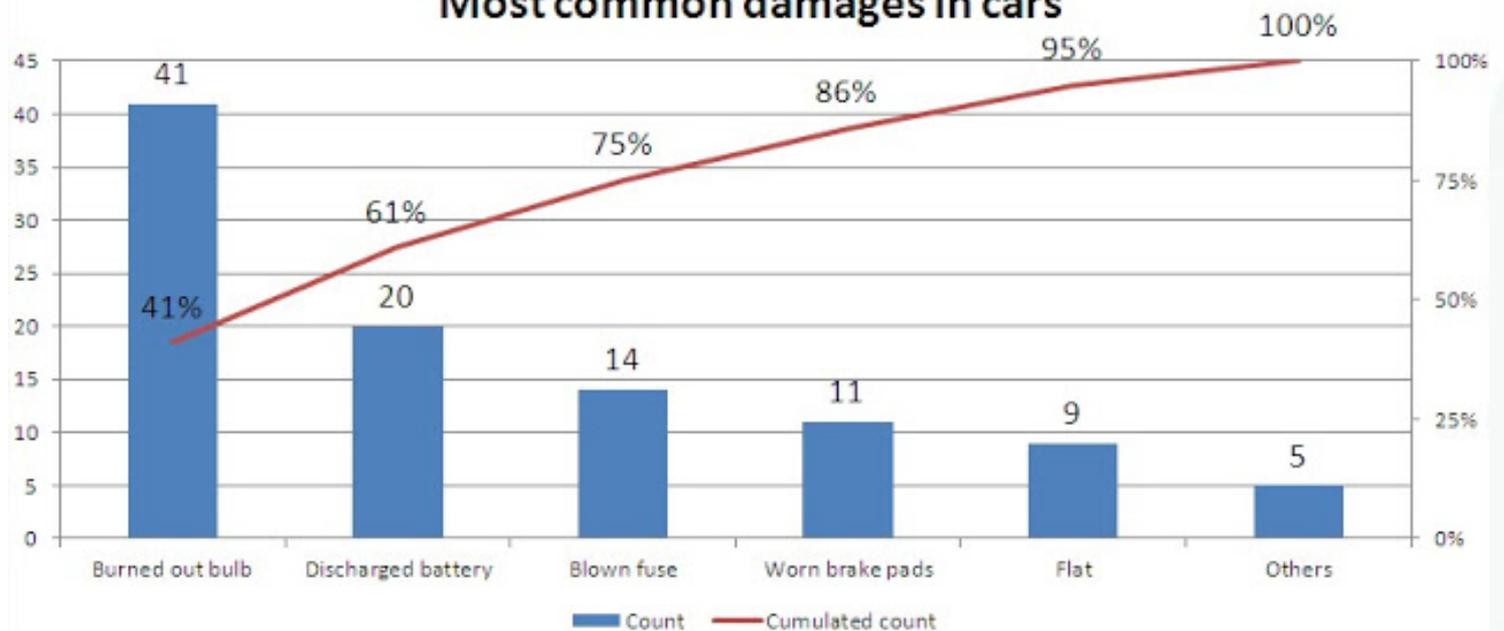
1. Define the problem to be analyzed
  - a) Identify potential causes
2. Determine criterion to use when comparing the possible causes,
  - a) Occurrence, consequences, costs
3. Define the time interval during which data will be collected and carry out the data collection for the selected criterion.
  - a) Often, data already available

# Pareto Chart – How To

4. Place the causes from left to right on the horizontal axis of the chart,
  - a) in descending relative importance.
  - b) Draw rectangles to heights that represent this importance.
5. Mark the data value on the left vertical axis and the percentage value on the right,
  - a) draw a curve of cumulative importance along the top edges of the rectangles.

## Pareto Diagram

### Most common damages in cars



# *Pareto Charts*

**EXERCISE**

# SCATTER PLOT (DIAGRAM)

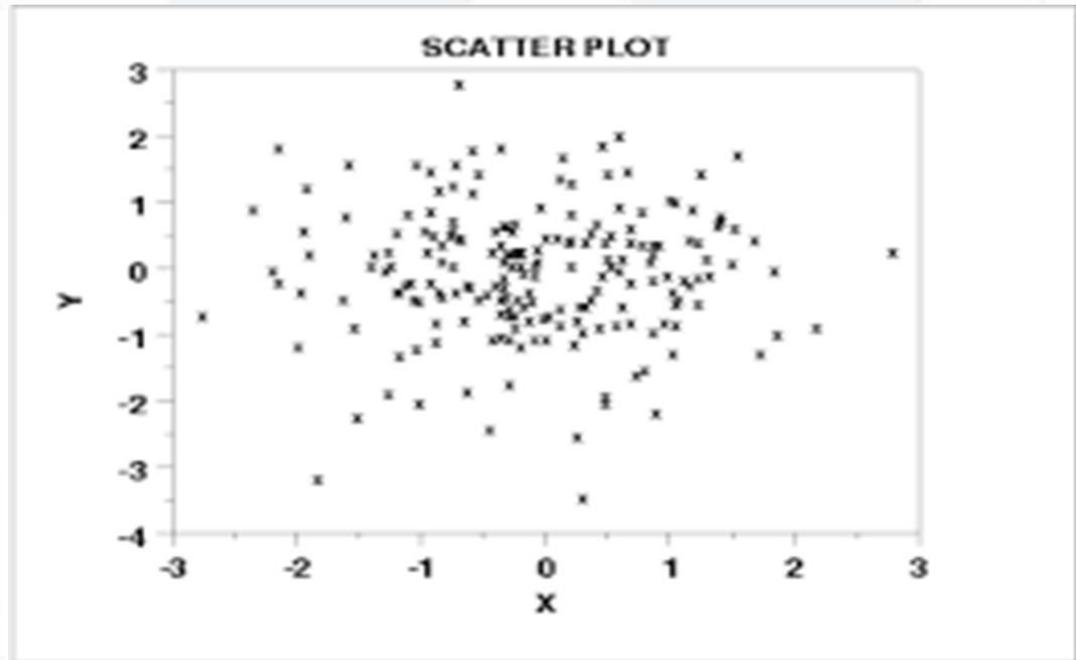
- Graphic display of data points useful for determining relationships between two variables
  - Independent variable (inputs)
  - Dependent variable (outputs)
- Shapes or trends can provide clues
- Useful for
  - Exploring impact one cause may have on another
  - Ruling out causes at different levels not linked to root cause

# SCATTER PLOT (DIAGRAM)

- Graphic display of data points useful for determining relationships between two variables
- Types of correlation
  - None (random)
  - Negative / Positive ; Linear / Non-Linear
  - Clusters
  - Cyclic

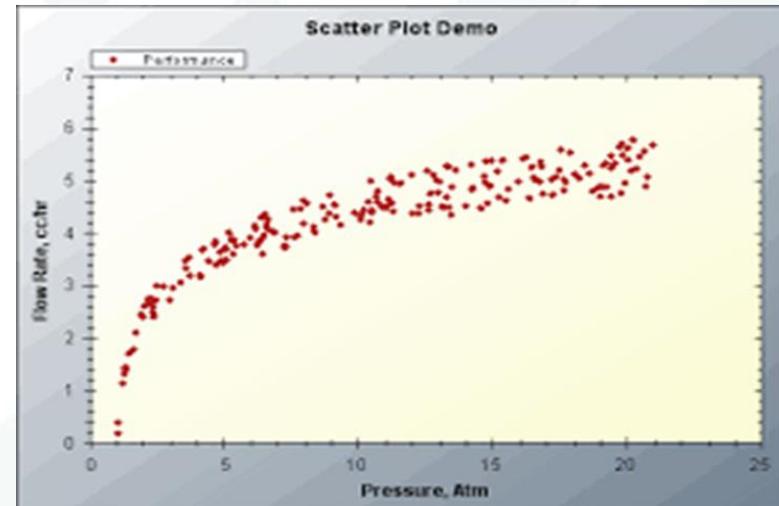
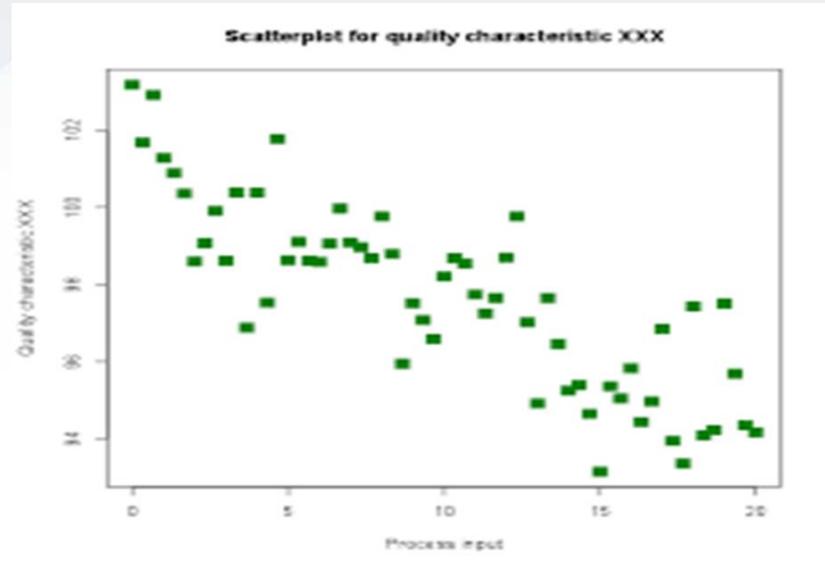
# SCATTER PLOT (DIAGRAM)

- Types of correlation
  - None (random)
  - Negative / Positive  
Linear / Non-Linear
  - Clusters
  - Cyclic



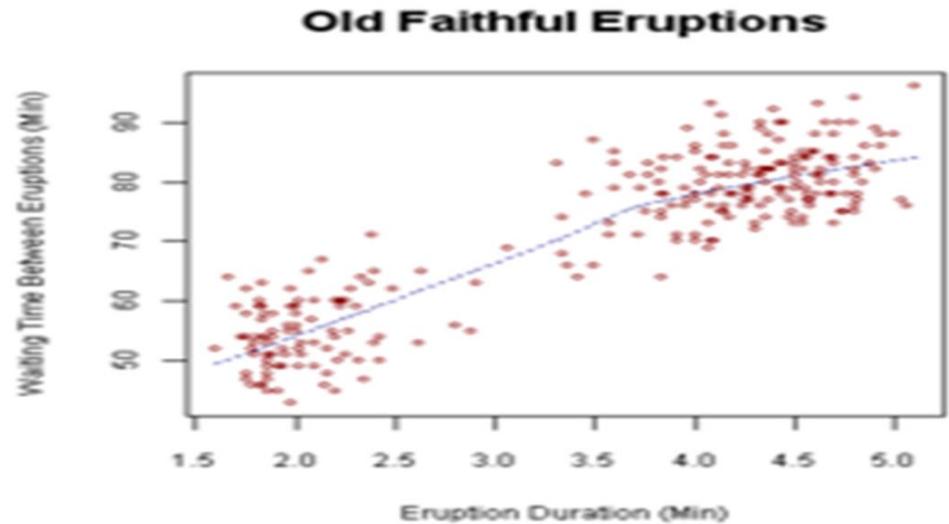
# SCATTER PLOT (DIAGRAM)

- Types of correlation
  - None (random)
  - Negative / Positive
  - Linear/ Non-linear
  - Clusters
  - Cyclic



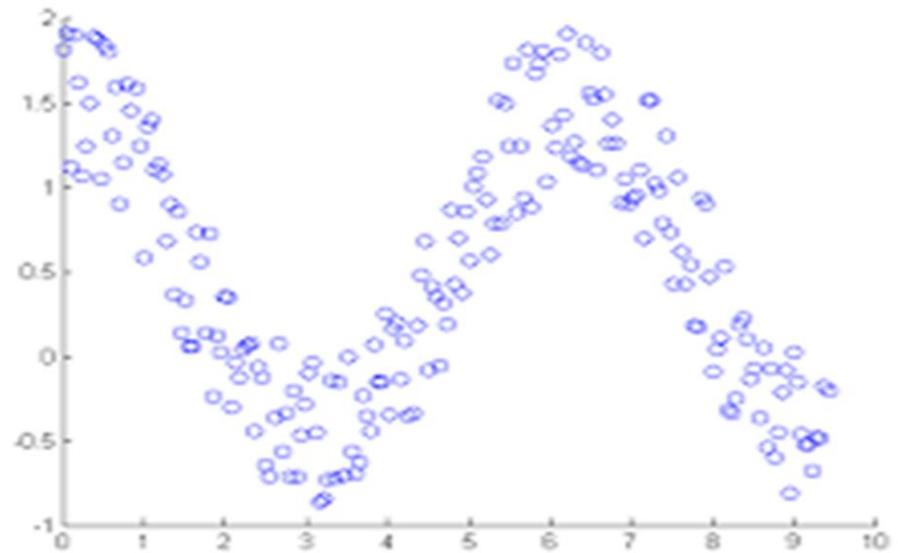
# SCATTER PLOT (DIAGRAM)

- Types of correlation
  - None (random)
  - Negative / Positive
  - Linear/ Non-linear
  - **Clusters**
  - Cyclic



# SCATTER PLOT (DIAGRAM)

- Types of correlation
  - None (random)
  - Negative / Positive
  - Linear/ Non-linear
  - Clusters
  - **Cyclic**



# Scatter Charts – How To

1. Select the two variables to be examined.
  - a) Independent (Input)
  - b) Dependent (Output)

*Independent variable - stands alone, is not influenced by other variable  
example - fertilizer*

*Dependent variable - changes based on other variables  
example – plant growth*

# Scatter Charts – How To

1. Select the two variables to be examined.
2. For each value of the independent variable (input), measure the corresponding value of the dependent variable (output)
3. X-axis will be independent variable; Y-axis will be dependent variable
4. Plot and analyze the collected data pairs in the chart
  - a) If no correlation found, also try a logarithmic chart which may reveal a trend not readily visible

# Scatter Plots

EXERCISE