Describing Through-Time Data: The Run Chart

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

Learning objective:

Construct a run chart using RStudio

Statistical Analysis

Statistical analysis has two parts:

- Graphics: pictures that provide a visual representation of what the numbers describe or identify

Statistical Analysis

- Numerics: numbers and statistical calculations which summarize and describe our data

Statistical Analysis

We always use both pictures and numbers (‘never present a picture without stats; never present stats without a picture’!)

Arranging and Presenting Data

The first step in the analysis and interpretation of data from a random sample is the arrangement and presentation of the data.

This should be done by first graphically describing the data.

Common Methods of Graphically Describing Sample Data

- Run Charts
- Frequency Distributions
  - ✓ Ungrouped
  - ✓ Grouped
  - ✓ Relative

Common Methods of Graphically Describing Sample Data

- Histograms
- Frequency Polygons
- Box and Whisker Plots

Presenting Data As Observed Through Time: Run Charts

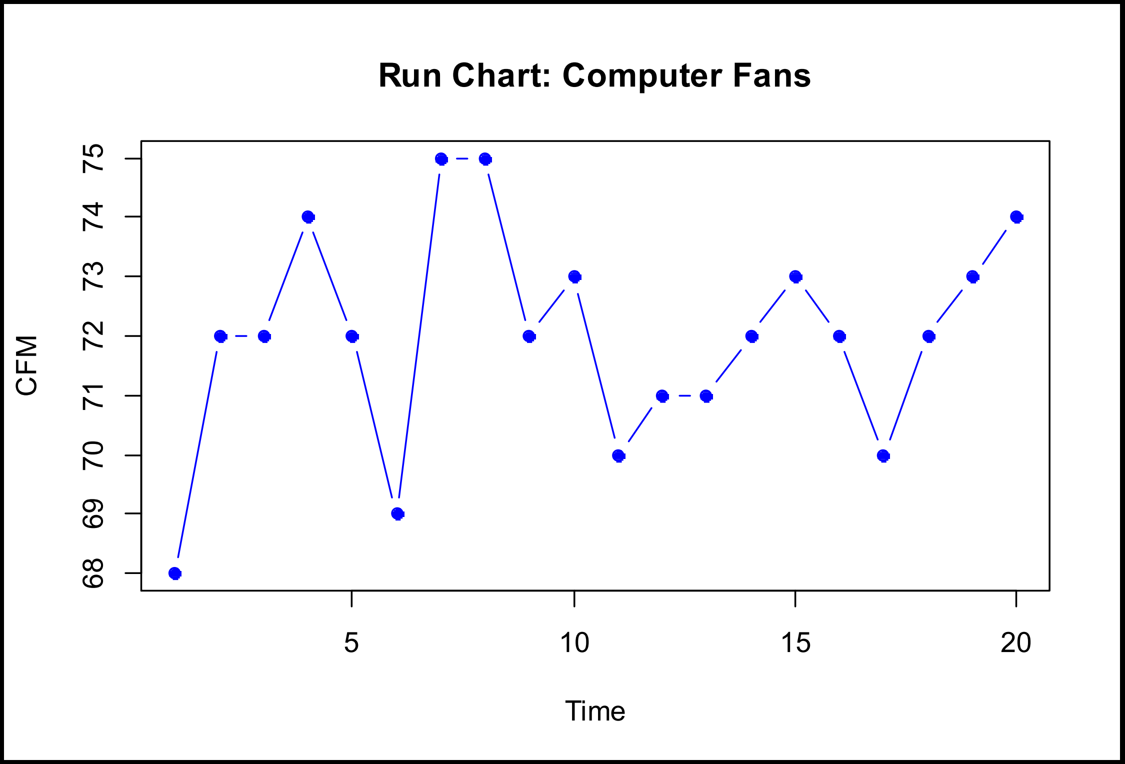
An engineer gathered 20 consecutive computer fans from a production line, keeping track of the order in which the fans were produced.

Presenting Data As Observed Through Time: Run Charts

Then these fans were tested for air flow in CFM. This testing produced the following data for the 20 fans, presented in time order.

Fans 1-10:	68	72	72	74	72	69	75	75	72	73
Fans 10-20:	70	71	71	72	73	72	70	72	73	74

Run Chart Example



Step 1: Create the Data File

Create a Vector

```
cfm <- c(68,72,72,74,72,69,75,75,72,73,70,71,71,72,73,72,70,72,73,74)
```

Store the Variable in a data frame

```
fans <- data.frame(cfm)
View(fans)
```

Step 2: Create the Run Chart

```
> require(lolcat)
> spc.run.chart(fans$cfm, main = "Run Chart: Computer Fans", ylab =
"CFM")
```

Step 3: Add a horizontal line

```
> abline(h=72)
```

Other Options for Customization

Point symbol: pch = (1-25)

Point size: cex =

Color: col = “red” (color name or hexadecimal code)

Line type: lty = (0-6)

Line width: lwd =

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Frequency Distributions

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

Learning objectives:

Construct an ungrouped frequency distribution using RStudio

Construct a grouped frequency distribution using RStudio

Frequency Distributions

Frequency distributions provide us with a method for arranging and viewing data sets. This allows for easier interpretation and analysis of the data.

Ungrouped vs Grouped Frequency Distributions

Use ungrouped when there are fewer than 20 unique data values in the data set

Use grouped when there are more than 20 unique data values in the data set

Ungrouped Frequency Distribution

Using the same fan data as we employed for the run chart:

Fans 1-10:	68	72	72	74	72	69	75	75	72	73
Fans 10-20:	70	71	71	72	73	72	70	72	73	74

Ungrouped Frequency Distribution Example

	value	freq	rel.freq	cum.up	cum.down
1	68	1	0.05	0.05	1.00
2	69	1	0.05	0.10	0.95
3	70	2	0.10	0.20	0.90
4	71	2	0.10	0.30	0.80
5	72	7	0.35	0.65	0.70
6	73	3	0.15	0.80	0.35
7	74	2	0.10	0.90	0.20
8	75	2	0.10	1.00	0.10

Where:

- value = Score, Value, or Observation
- freq = Frequency
- rel.freq = Relative Frequency
- cum.up / cum.down = Cumulative

Ungrouped Frequency Distribution Example

	value	freq	rel.freq	cum.up	cum.down
1	68	1	0.05	0.05	1.00
2	69	1	0.05	0.10	0.95
3	70	2	0.10	0.20	0.90
4	71	2	0.10	0.30	0.80
5	72	7	0.35	0.65	0.70
6	73	3	0.15	0.80	0.35
7	74	2	0.10	0.90	0.20
8	75	2	0.10	1.00	0.10

Frequency distributions are considered ‘ungrouped’ when each row, or ‘class interval’, consists of only one score, value, or observation.

Ungrouped Frequency Distribution in R

```
> frequency.dist.ungrouped(fans$cfm)
```

Grouped Frequency Distribution

Ungrouped frequency distributions have one value for each class interval. Where the Range ($X_H - X_L$) of the data set is large, however, constructing a functional ungrouped frequency distribution becomes untenable.

Grouped Frequency Distribution

In these cases, we use a Grouped Frequency Distribution.

Grouped frequency distributions have a range of values associated with each interval.

- Example interval: 5 – 9
- Example interval: 1.230 - 1.234

Grouped Frequency Distribution Example

Forty (40) castings for use in a machining process have been randomly selected from an incoming lot from a supplier.

Grouped Frequency Distribution Example

Descriptive Statistics

Variable	Sample Size (n)	Mean	Std. Dev.	Low	High	Range
Weight	40	134.75	14.75	109	170	61

The data are initially arranged in an ungrouped frequency distribution:

Ungrouped Frequency Distribution

Too Many Intervals

	value	freq	rel.freq	cum.up	cum.down
1	109	1	0.025	0.025	1.000
2	111	1	0.025	0.050	0.975
3	117	1	0.025	0.075	0.950
4	118	1	0.025	0.100	0.925
5	120	1	0.025	0.125	0.900
6	121	1	0.025	0.150	0.875
7	122	2	0.050	0.200	0.850
8	124	2	0.050	0.250	0.800
9	125	1	0.025	0.275	0.750
10	126	2	0.050	0.325	0.725
11	128	2	0.050	0.375	0.675
12	129	3	0.075	0.450	0.625
13	130	1	0.025	0.475	0.550
14	131	2	0.050	0.525	0.525
15	132	1	0.025	0.550	0.475
16	133	1	0.025	0.575	0.450
17	134	1	0.025	0.600	0.425
18	135	2	0.050	0.650	0.400
19	137	1	0.025	0.675	0.350
20	139	1	0.025	0.700	0.325
21	143	2	0.050	0.750	0.300
22	146	1	0.025	0.775	0.250
23	148	2	0.050	0.825	0.225
24	152	1	0.025	0.850	0.175
25	155	2	0.050	0.900	0.150
26	158	1	0.025	0.925	0.100
27	162	1	0.025	0.950	0.075
28	165	1	0.025	0.975	0.050
29	170	1	0.025	1.000	0.025

Grouped Frequency Distribution

The data are then reorganized in a Grouped Frequency distribution

	l	min	midpoint	max	u	freq	rel.freq	cum.up	cum.down
1	[	105	107.5	110	)	1	0.025	0.025	1.000
2	[	110	112.5	115	)	1	0.025	0.050	0.975
3	[	115	117.5	120	)	2	0.050	0.100	0.950
4	[	120	122.5	125	)	6	0.150	0.250	0.900
5	[	125	127.5	130	)	8	0.200	0.450	0.750
6	[	130	132.5	135	)	6	0.150	0.600	0.550
7	[	135	137.5	140	)	4	0.100	0.700	0.400
8	[	140	142.5	145	)	2	0.050	0.750	0.300
9	[	145	147.5	150	)	3	0.075	0.825	0.250
10	[	150	152.5	155	)	1	0.025	0.850	0.175
11	[	155	157.5	160	)	3	0.075	0.925	0.150
12	[	160	162.5	165	)	1	0.025	0.950	0.075
13	[	165	167.5	170	)	1	0.025	0.975	0.050
14	[	170	172.5	175	)	1	0.025	1.000	0.025

Grouped Frequency Distribution in R

```
> frequency.dist.grouped(castings$weight)
```

Grouped Frequency Distribution

Important questions to answer:

- How many class intervals, optimally, should the frequency distribution have? How many is too few? Too many?

Grouped Frequency Distribution

- What class interval size is best for the data set we are attempting to portray in a frequency distribution?
- At what class interval should we start the grouped frequency distribution?

Constructing a Grouped Frequency Distribution

- Generate a frequency distribution with as close as you can get to 10 class intervals, without going under (divide the Range by 10 for an estimate of the class interval size you'll need);

Constructing a Grouped Frequency Distribution

- Use one of the following class interval sizes: 1, 2, 3, or 5; increasing the sizes in multiples of 10 where required (e.g. 10, 20, 30, 50, 100…)

Constructing a Grouped Frequency Distribution

- Start the first class interval with a number that is a multiple of the class interval size
- The first class interval must contain the lowest score in the data set (X_L)

Constructing a Grouped Frequency Distribution

- lolcat::freq.dist.grouped considers all of these rules to give an optimal result

Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Frequency Polygons and Histograms

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

Learning objectives:

Create a Frequency Polygon using RStudio
Create a histogram using RStudio

Frequency Polygons and Histograms

Useful for:

- Evaluating a manufacturing or business process
- Determining machine and process capabilities

Frequency Polygons and Histograms

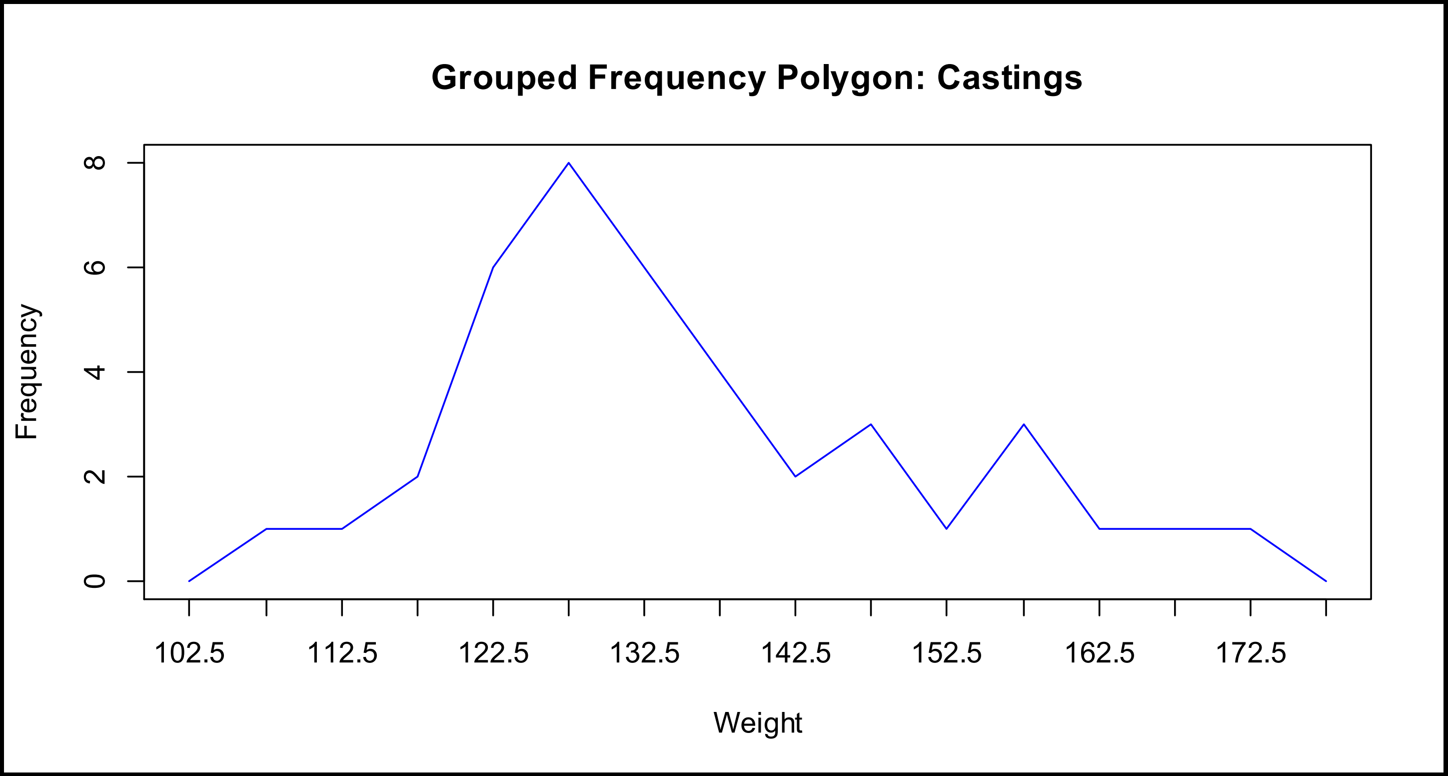
- Comparing material, vendor, operator, process and product characteristics

Ungrouped vs Grouped Frequency Histograms/Polygons

Use ungrouped when there are fewer than 20 unique data values in the data set

Use grouped when there are more than 20 unique data values in the data set

Frequency Polygons



Frequency Polygons

A graph or chart which represents the frequency of observations at each class interval (grouped) or value/score (ungrouped).

Similar to the frequency column of the frequency distribution.

Frequency Polygon: Advantages

Frequency polygons often present a more representative illustration of the data pattern when data are measured along a continuous scale.

Frequency Polygon: Advantages

The polygon becomes increasingly smooth and curve-like as the number of class intervals and sample size (n) increases, more closely representing the sampled population.

Ungrouped Frequency Polygon

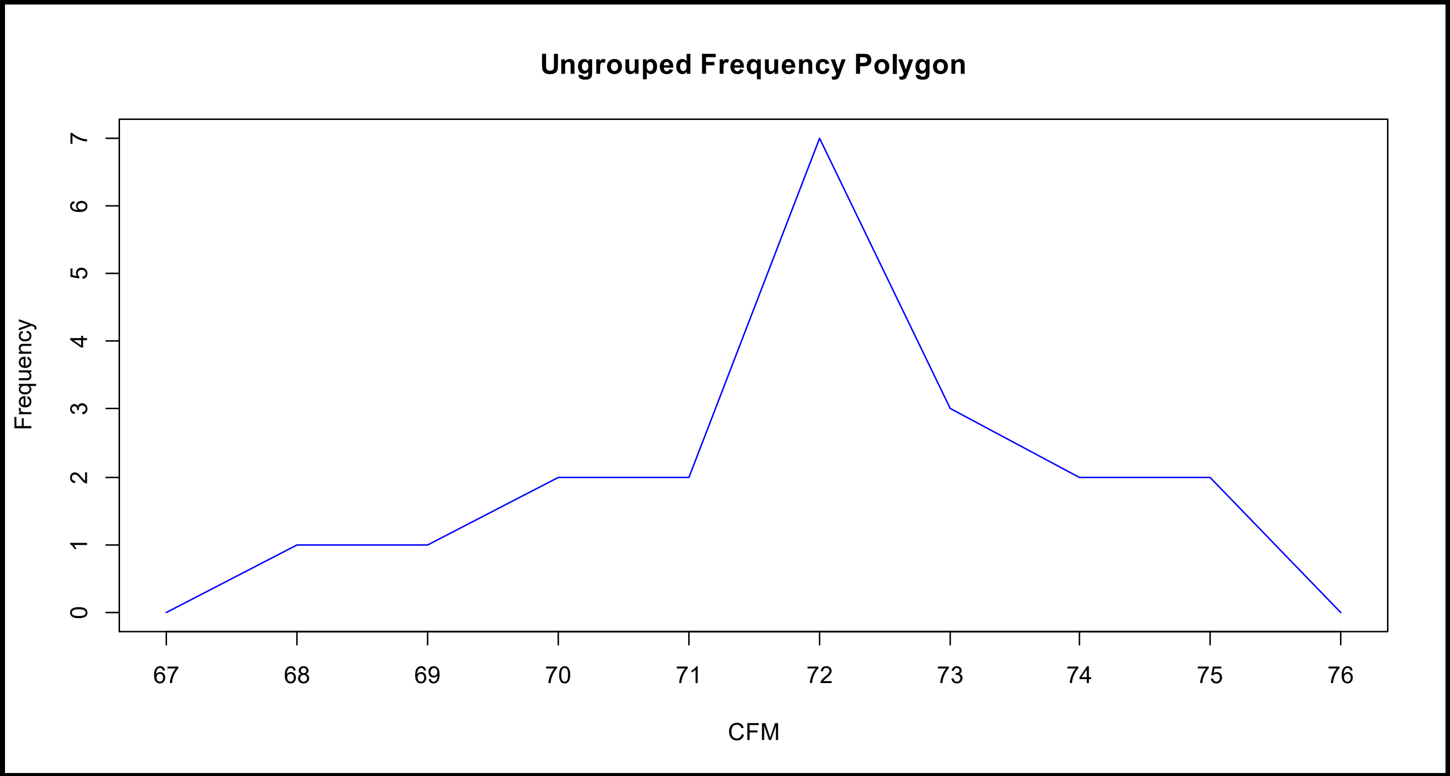
Using the same fan data as we employed for the ungrouped frequency distribution:

Fans 1-10:	68	72	72	74	72	69	75	75	72	73
Fans 10-20:	70	71	71	72	73	72	70	72	73	74

Ungrouped Frequency Polygon in R

```
> frequency.polygon.ungrouped(fans$cfm)
```

Ungrouped Frequency Polygon



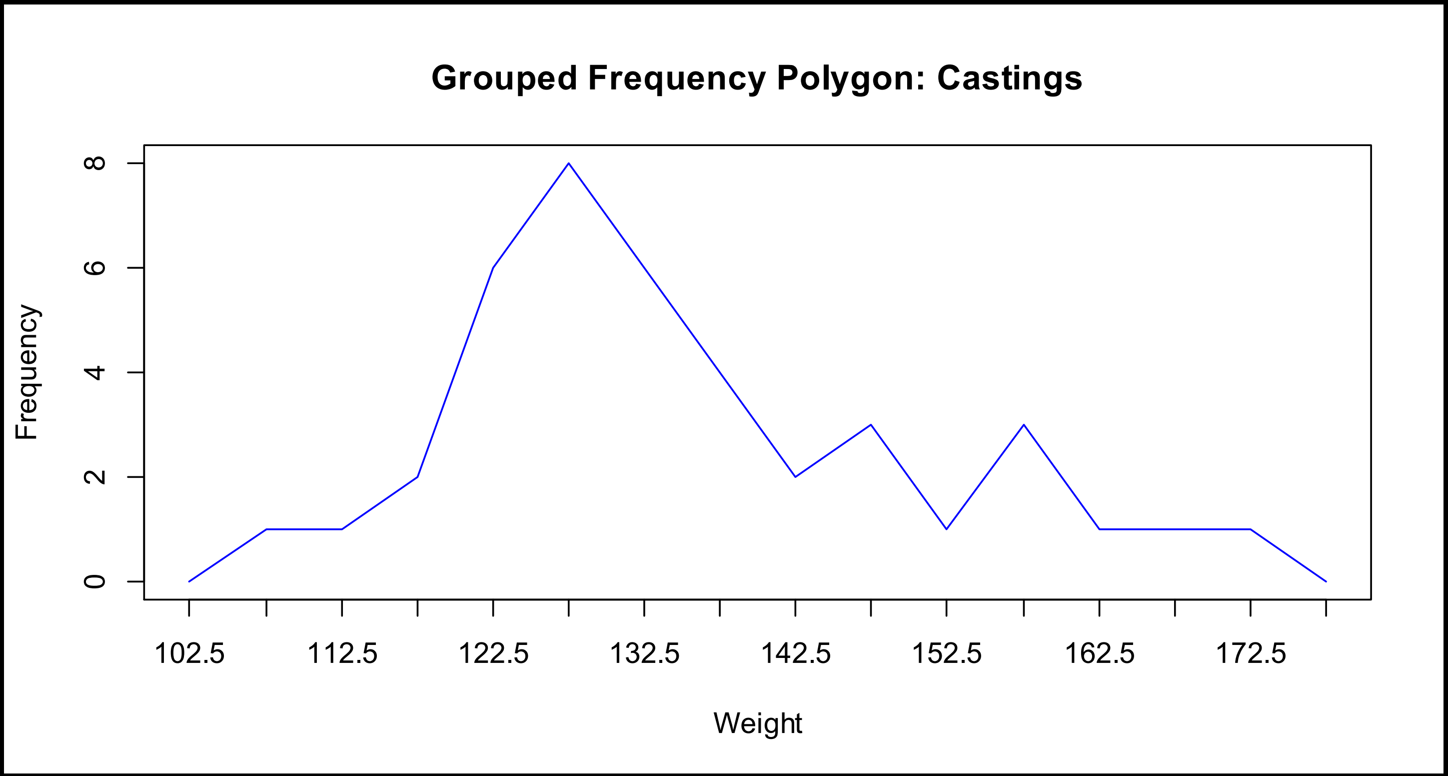
Grouped Frequency Polygon

Using the same castings data as we employed for the grouped frequency distribution:

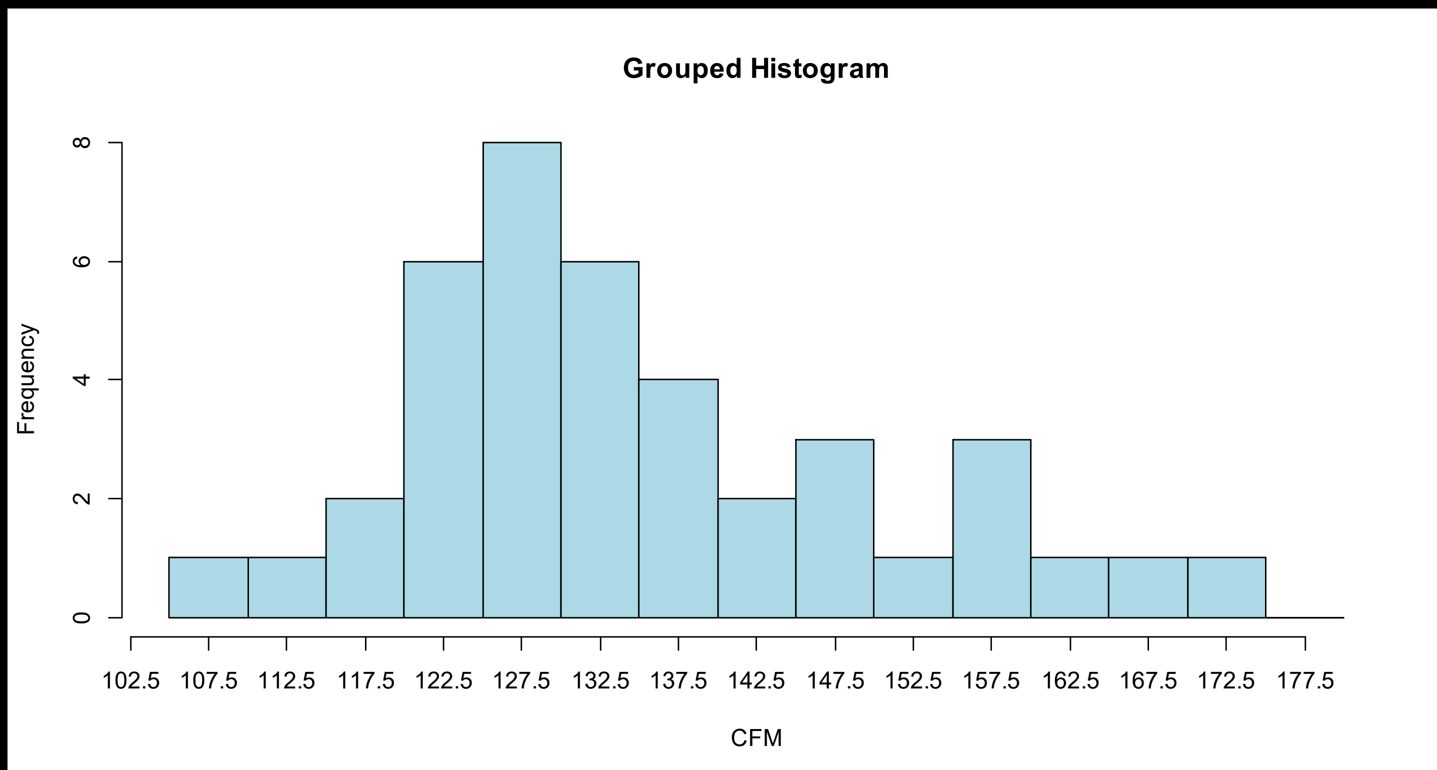
Grouped Frequency Polygon in R

```
> frequency.polygon.grouped(castings$weight)
```

Grouped Frequency Polygon



Histograms



Grouped Histogram
Frequency
0
2
4
6
8
102.5
107.5
112.5
117.5
122.5
127.5
132.5
137.5
142.5
147.5
152.5
157.5
162.5
167.5
172.5
177.5
CFM

Histograms

Similar to the frequency polygon, except that bars are used to represent the frequency of occurrence at each score or class interval.

Typically, each vertical bar in the histogram is centered above each class interval (or individual score).

Histogram: Advantages

Each bar or rectangular area clearly shows the relative magnitude of that class interval.

The area in each bar reflects the true proportion of the total number of observations occurring in the class interval.

A Note About Histograms

When the data represent discrete values, such as counts, histograms must be used.

When the data represent continuous values, a frequency polygon or histogram may be used.

Ungrouped Histogram

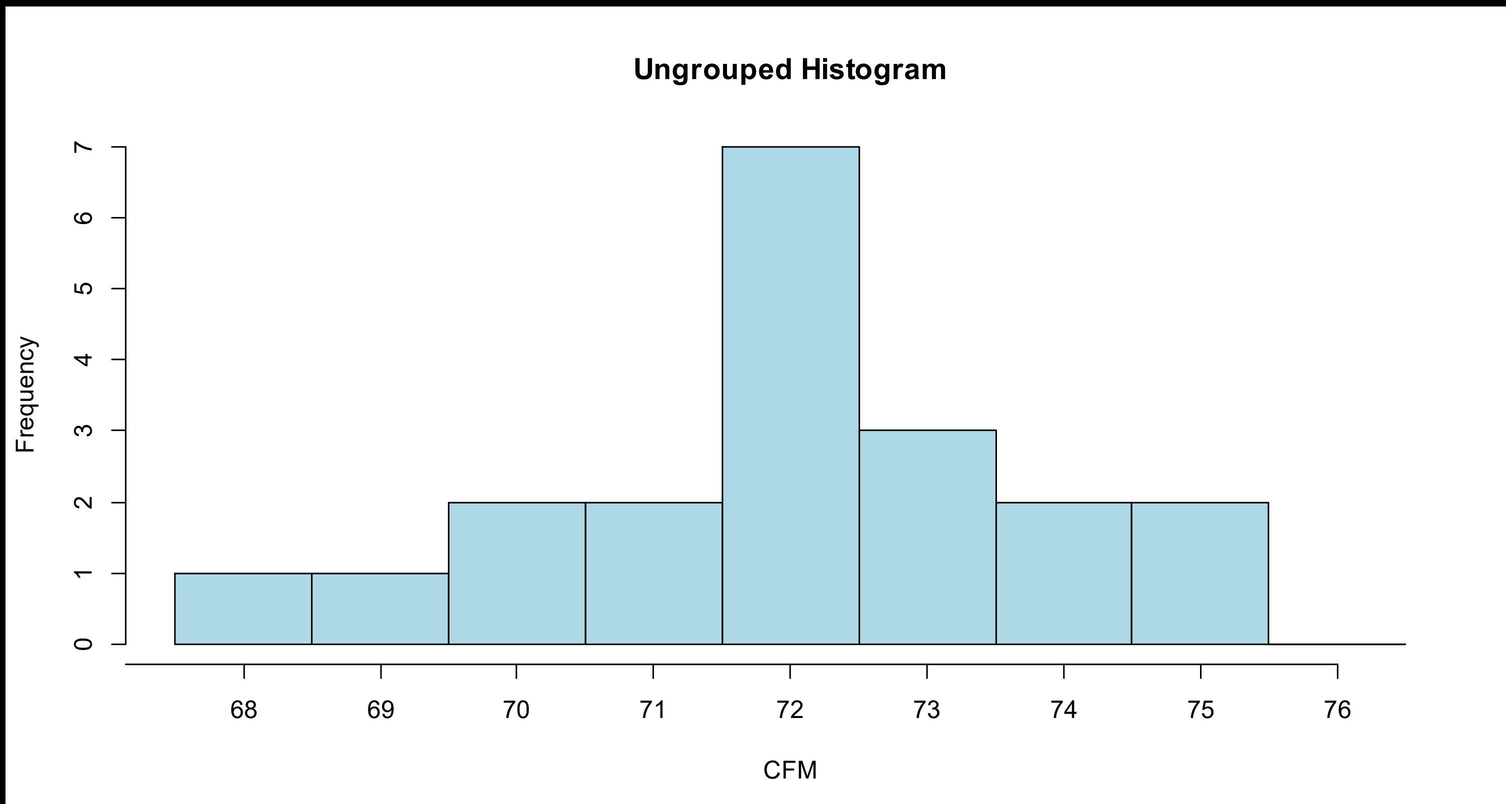
Using the same fan data as we employed for the ungrouped frequency polygon:

Fans 1-10:	68	72	72	74	72	69	75	75	72	73
Fans 10-20:	70	71	71	72	73	72	70	72	73	74

Ungrouped Histogram in R

```
> hist.ungrouped(fans$cfm)
```

Ungrouped Histogram



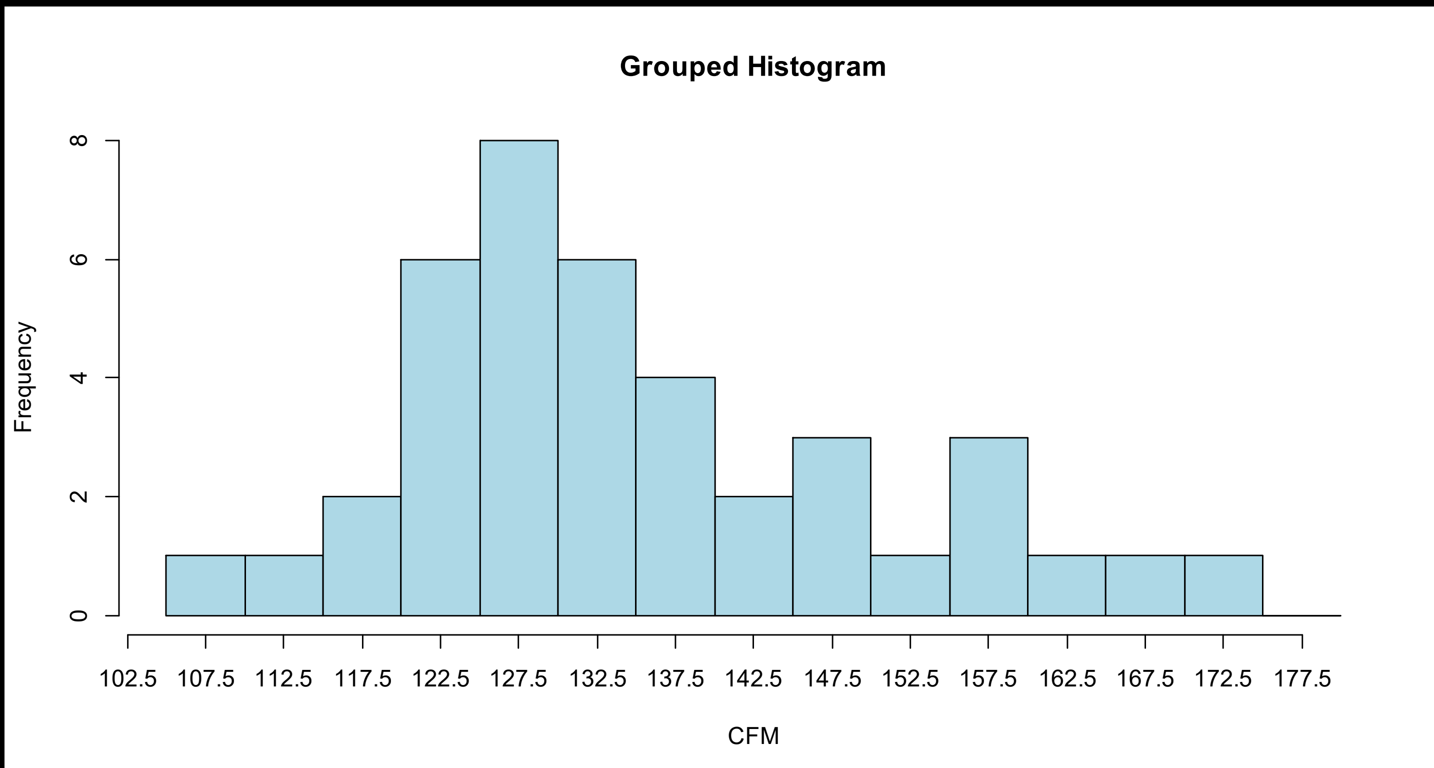
Grouped Histogram

Using the same castings data as we employed for the grouped frequency polygon:

Grouped Histogram in R

```
> hist.grouped(castings$weight)
```

Grouped Histogram



Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Histogram Patterns
Density Plots

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

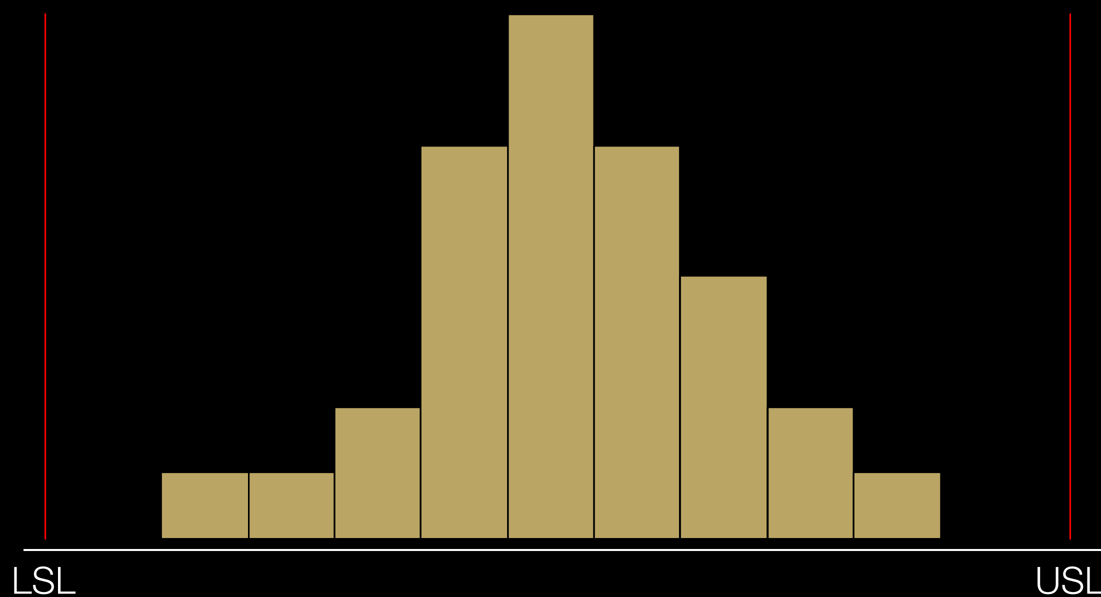
Learning objectives:

Interpret Histogram Patterns
Create a Density Plot using RStudio

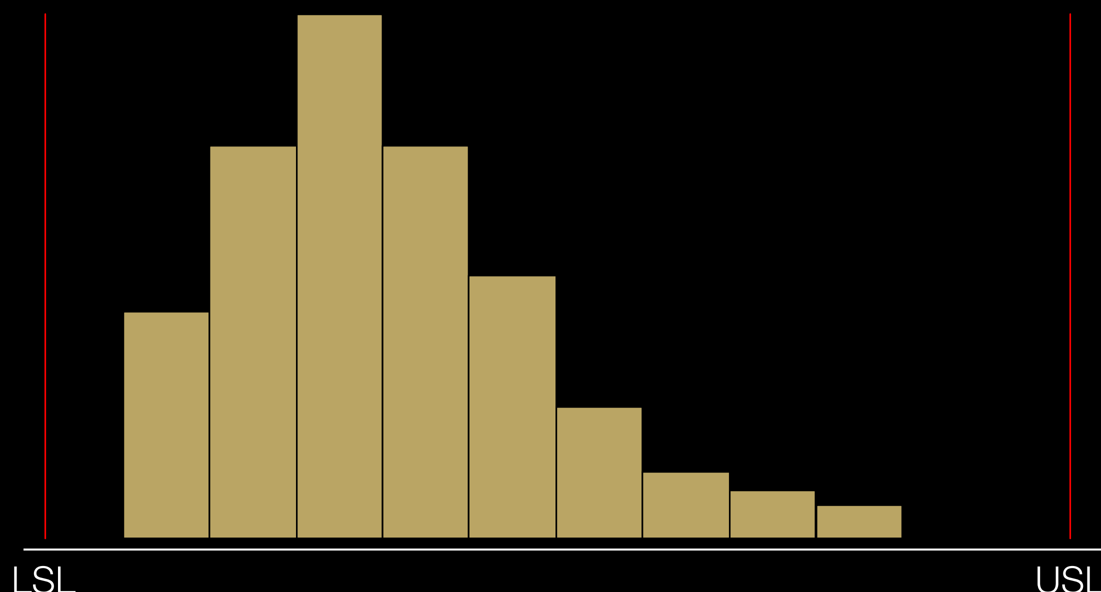
Histogram Patterns

The center, spread and shape of a histogram can give us clues as to what the data are telling us.

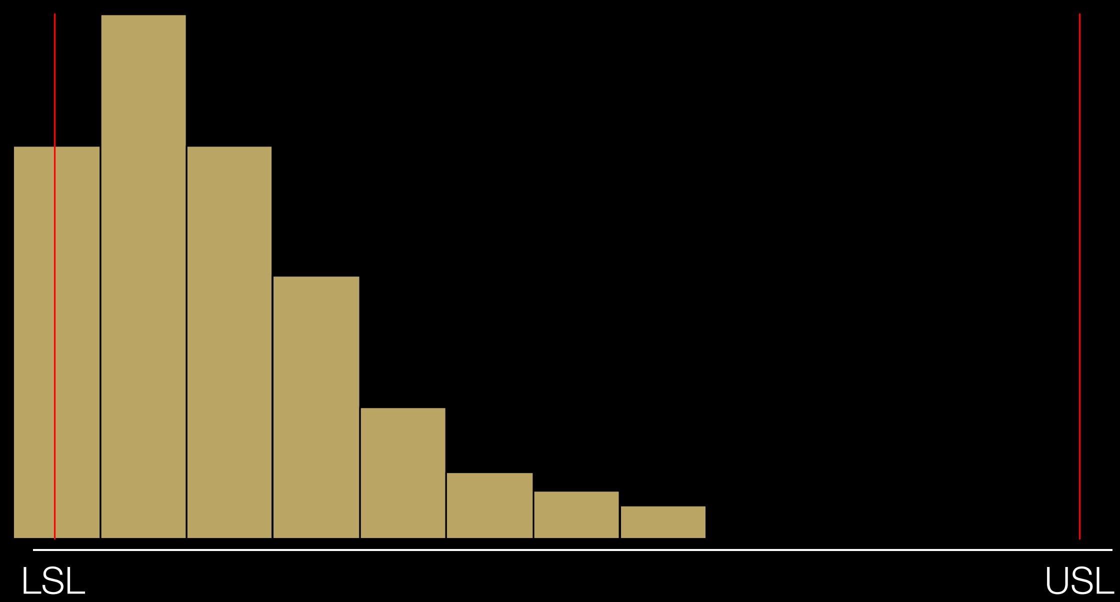
Pattern 1



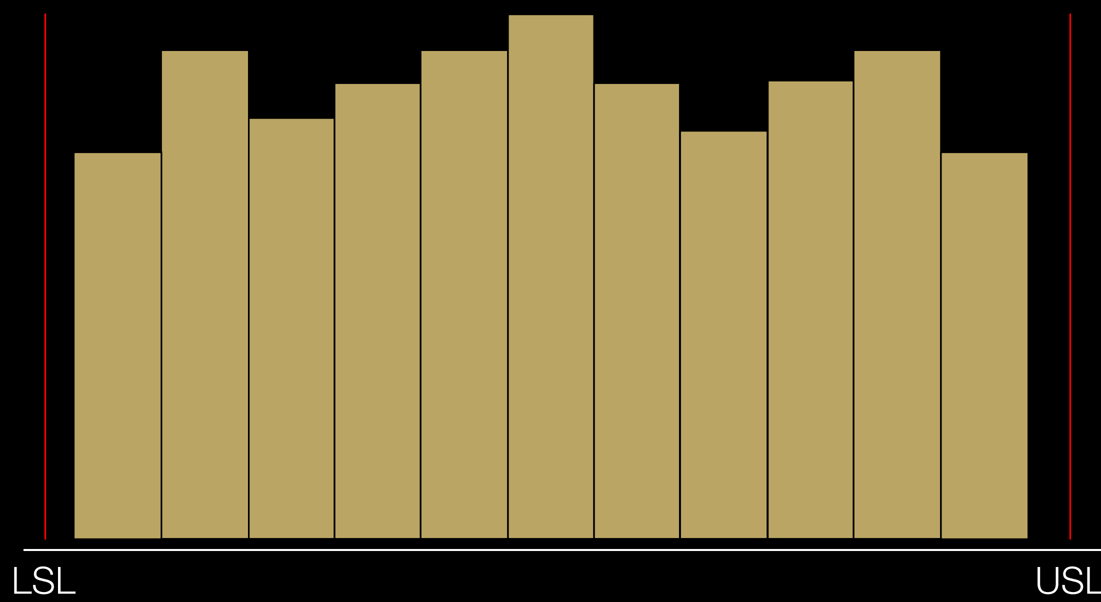
Pattern 2



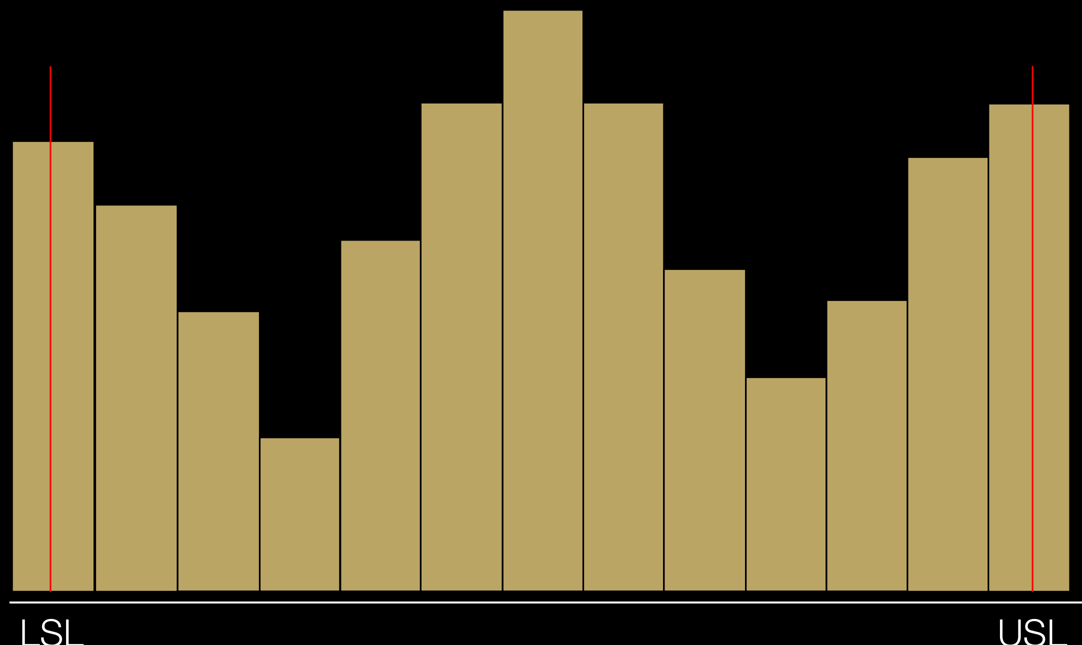
Pattern 3



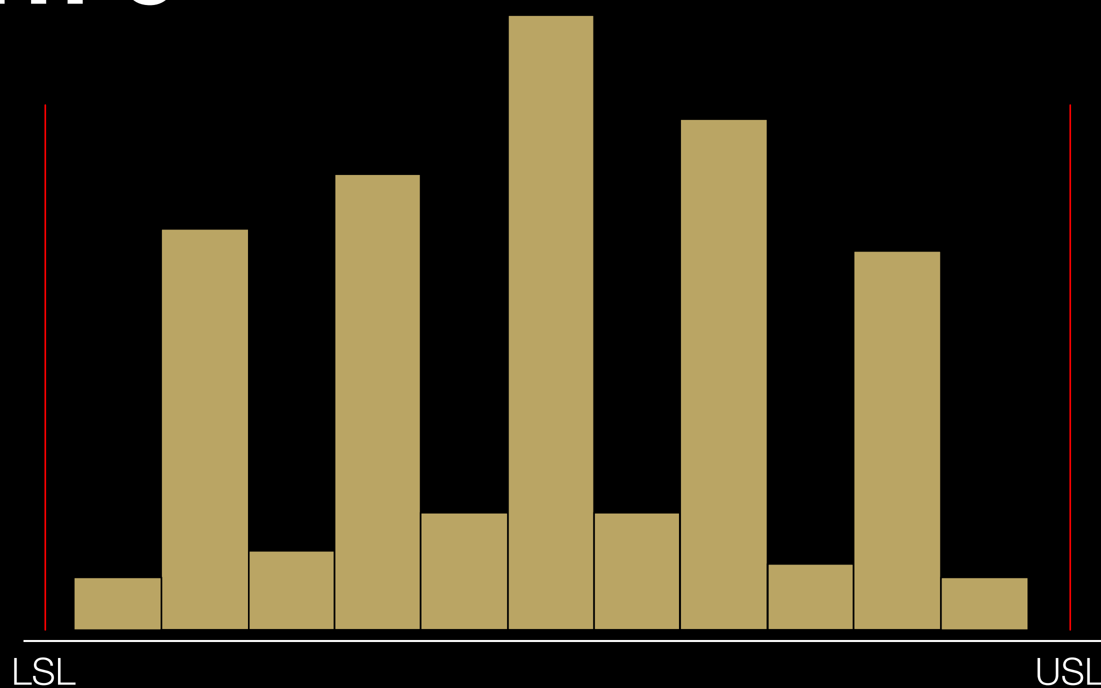
Pattern 4



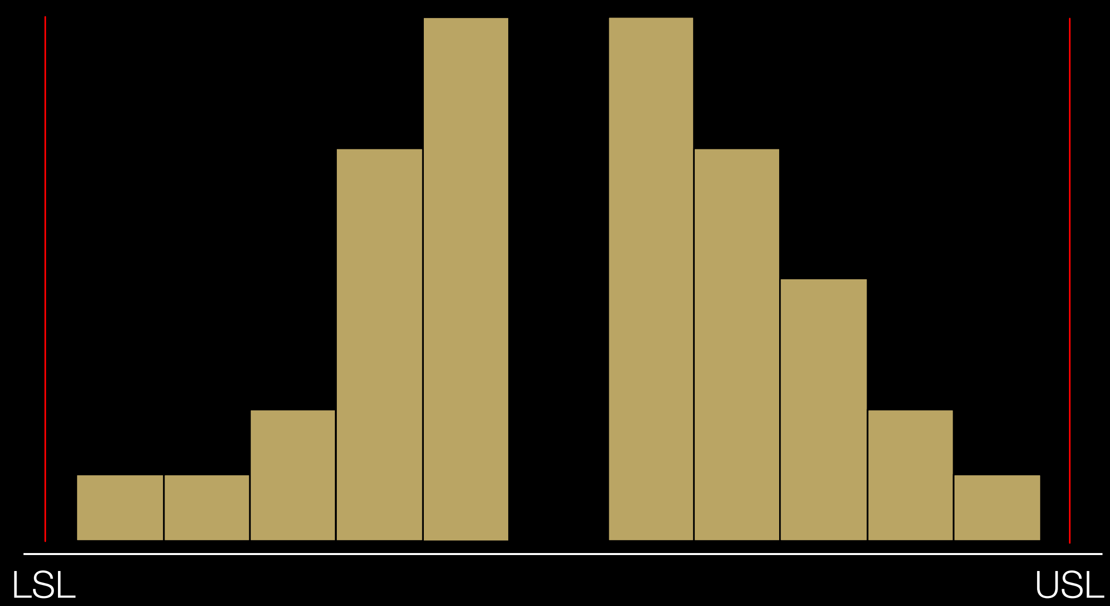
Pattern 5



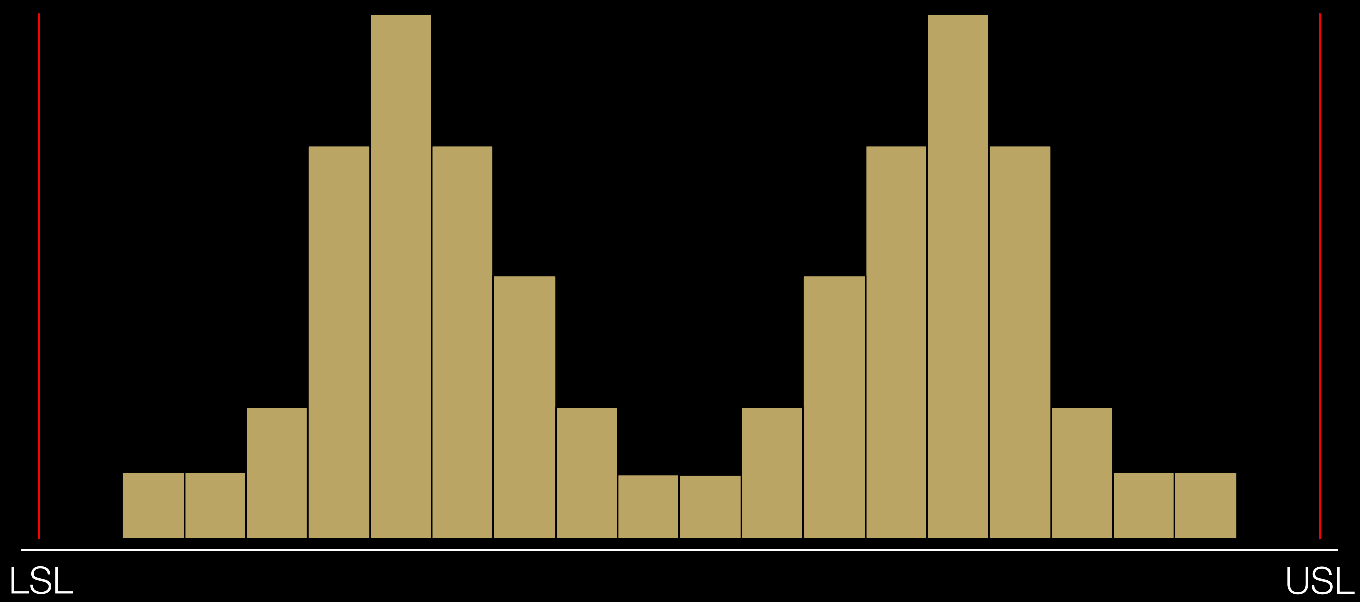
Pattern 6



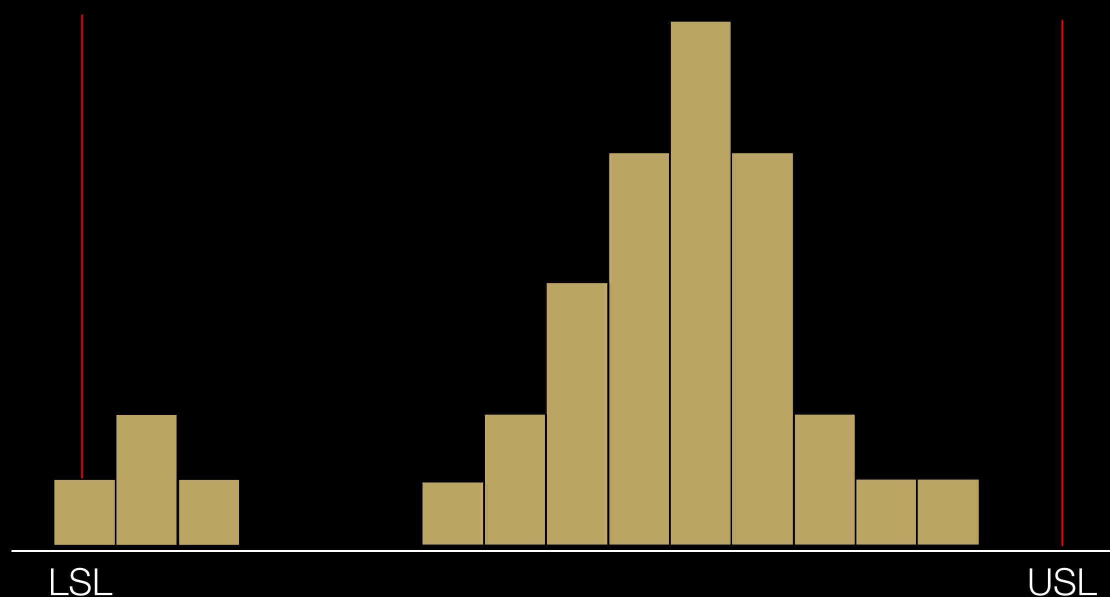
Pattern 7



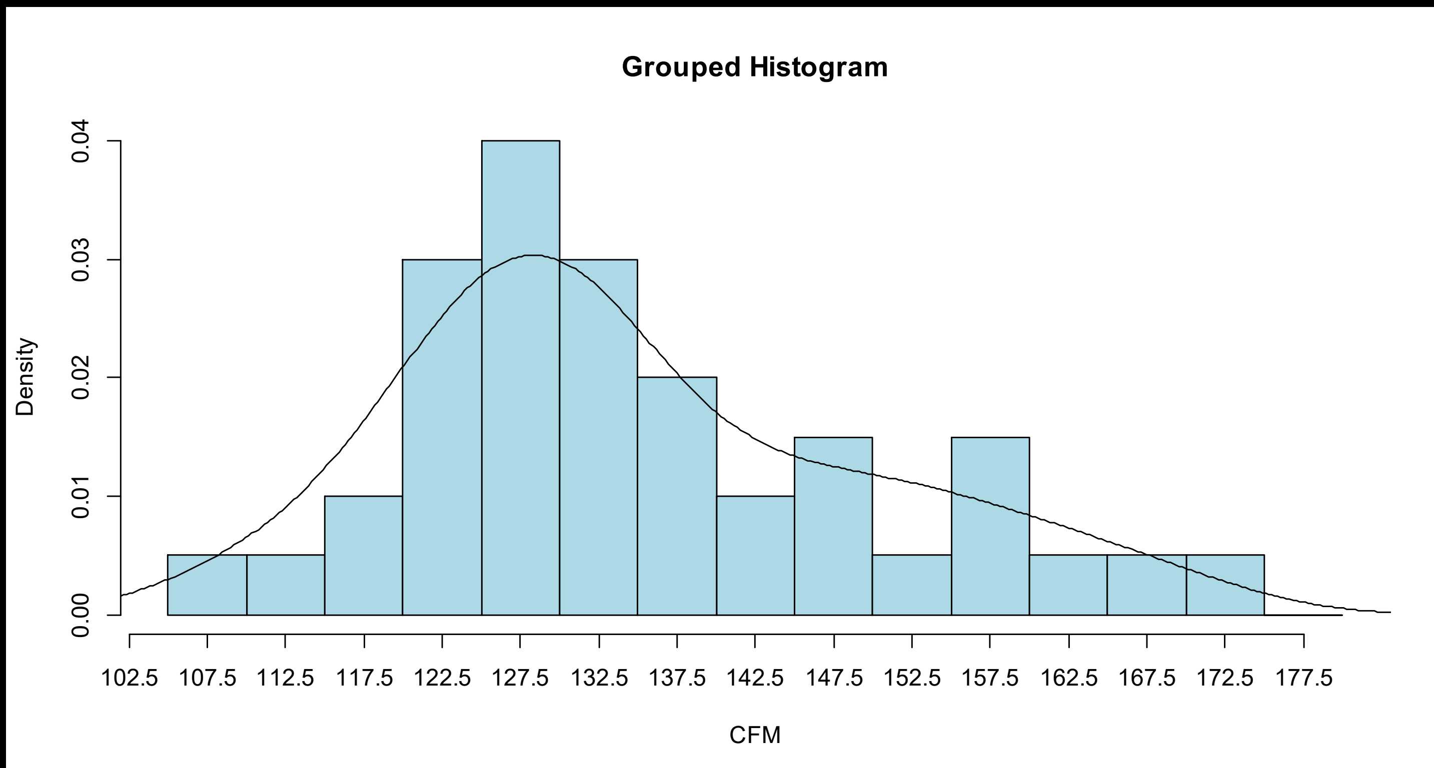
Pattern 8



Pattern 9



Density Plot



Density Plot

Similar to the frequency polygon, in that it is used with continuous data

Used to visualize an underlying probability distribution

Density Plot

When the data are continuous, we can use a density plot over a histogram.

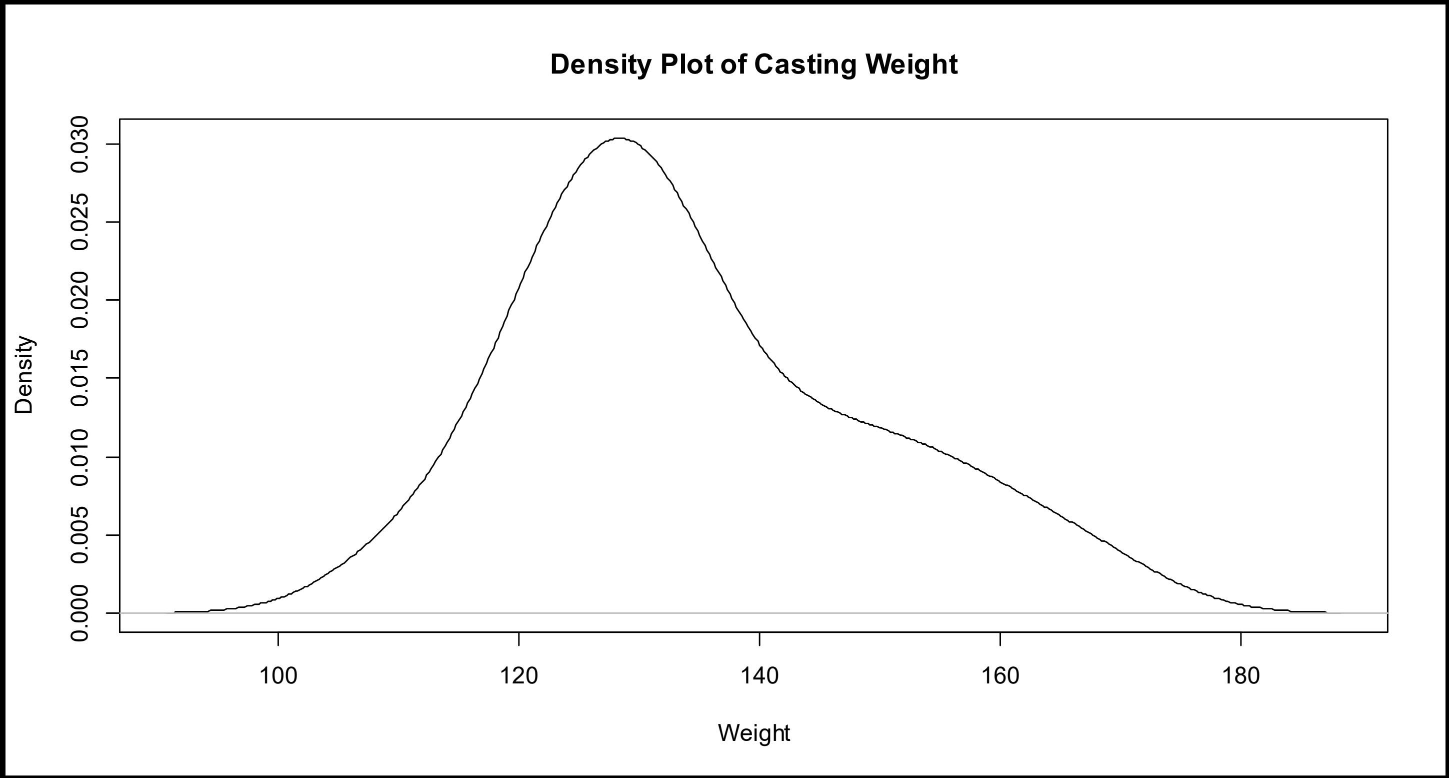
```
> hist.grouped(castings$weight, freq=F)
> lines(density(castings$weight))
```

Density Plot

The density plot can also be plotted without a histogram:

```
> plot(density(castings$weight)
```

Density Plot

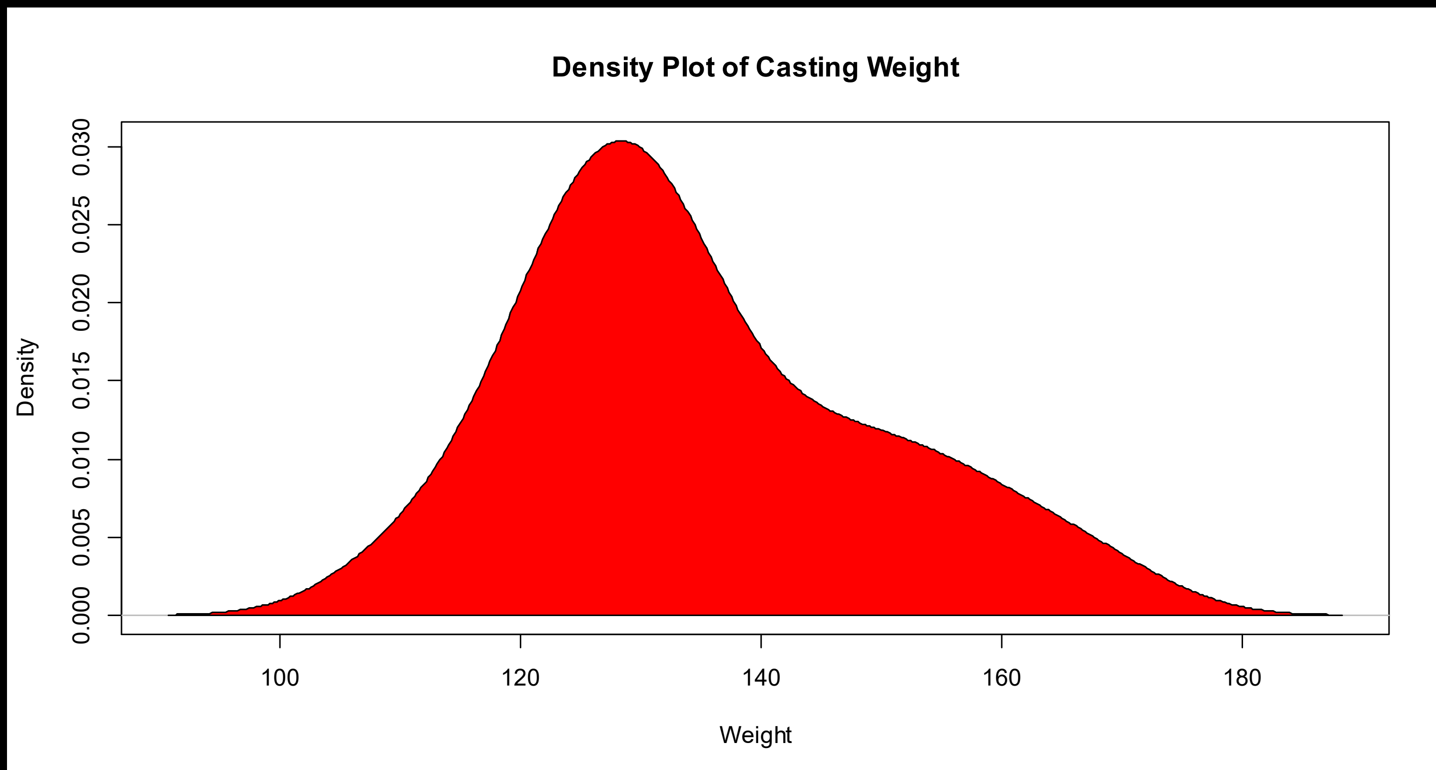


Density Plot

To fill a density plot with color:

```
> dp<-density(castings$weight)
> plot(dp, main="Density Plot of Casting
Weight", xlab="Weight")
> polygon(dp, col="red", border="black")
```

Density Plot



Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005

Box and Whisker Plots

Data Science for Quality Management: Describing Data Graphically

with **Wendy Martin**

Learning objective:

Create a Box and Whisker Plot using RStudio

Box and Whisker Plot

Box & Whisker Plots are used to display data corresponding to Percentiles, and typically from two or more sources or process streams, simultaneously

Box and Whisker Plot

One distinct advantages of this display is that the two sample data sets do not have to possess the same shape, but are directly comparable nonetheless.

Box and Whisker Plot

A second major advantage is that the Box & Whisker plot can display outliers; which we will see later can represent Special Causes of Variation.

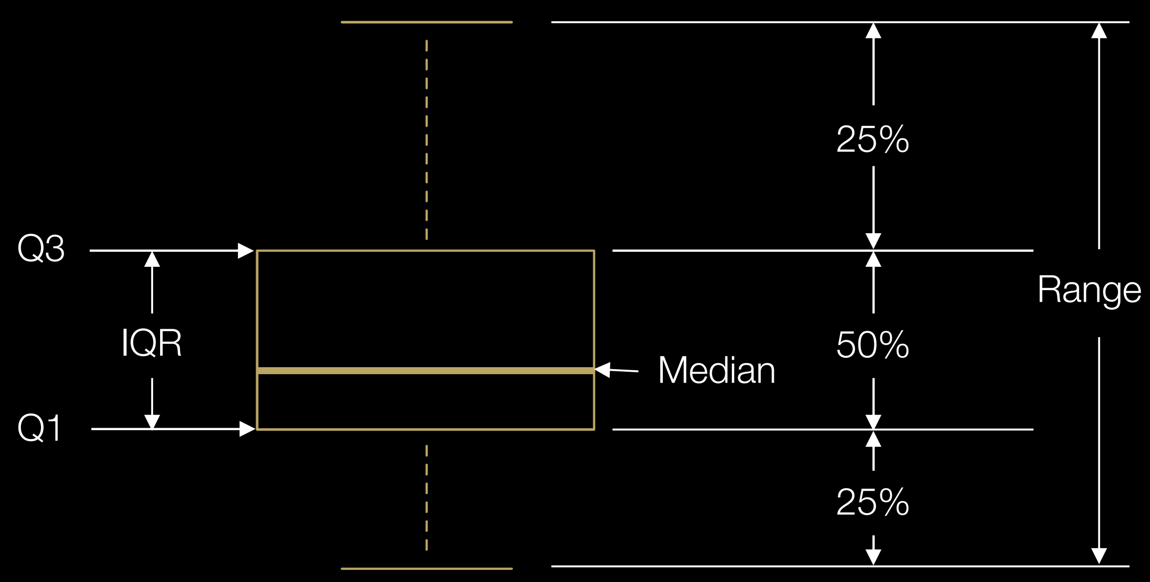
5 Number Summary

Maximum
Q3 (3rd Quartile)
Median (Q2) (2nd Quartile)
Q1 (1st Quartile)
Minimum

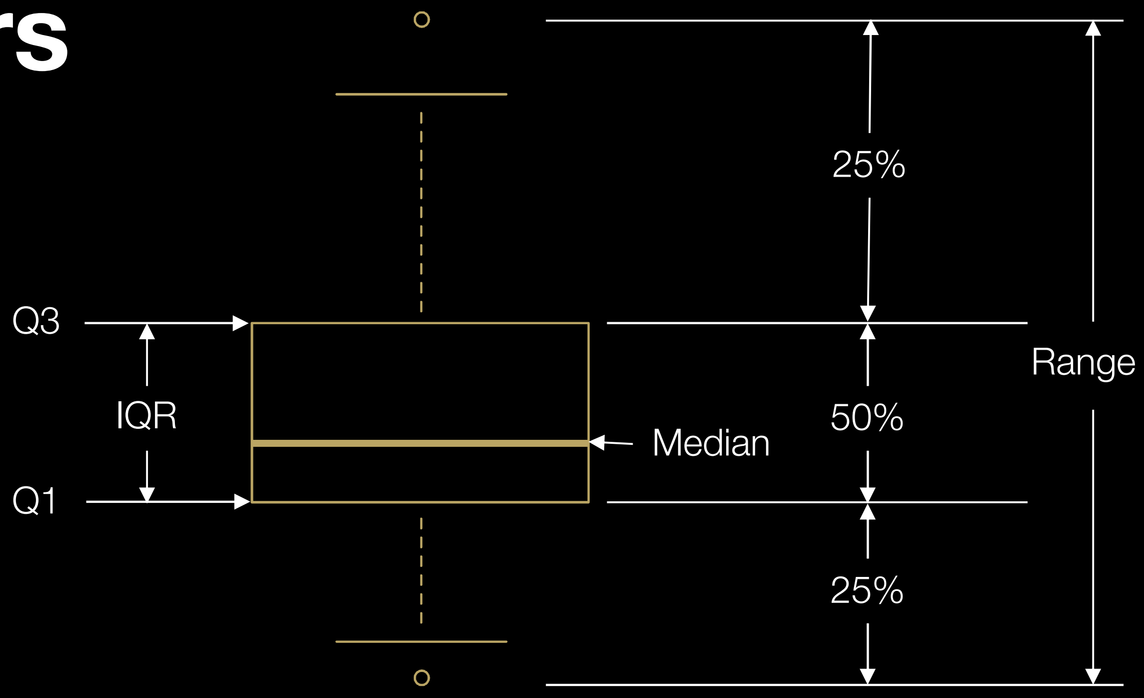
5 Number Summary

```
> summary(castings$weight)
```

Box and Whisker Plot



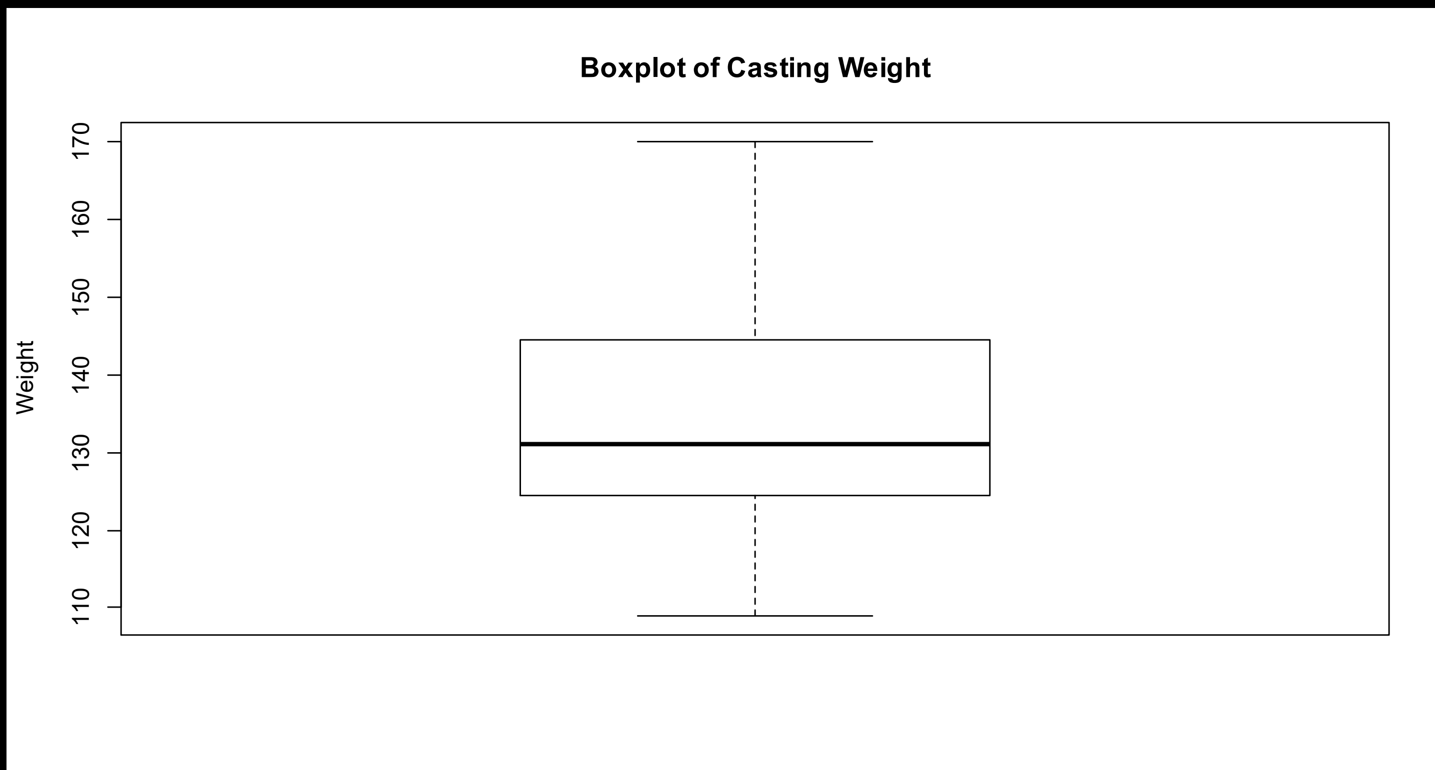
Box and Whisker Plot with Outliers



Box and Whisker Plot in R

```
> boxplot(castings$weight)
```

Box and Whisker Plot Example

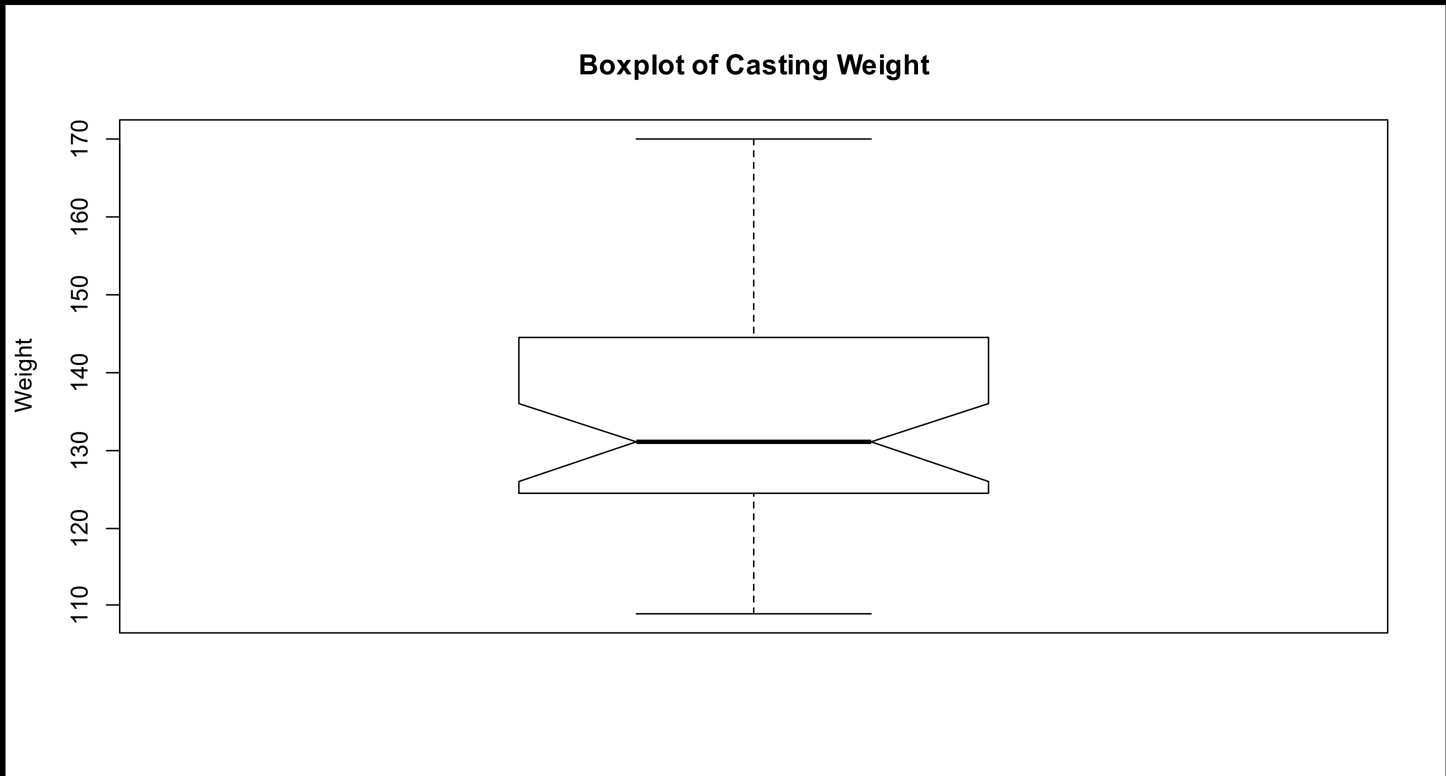


Notched Box and Whisker Plot

A notched Box and Whisker plot shows the 95% confidence interval of the median.

```
> boxplot(castings$weight, notch=T)
```

Notched Box and Whisker Plot



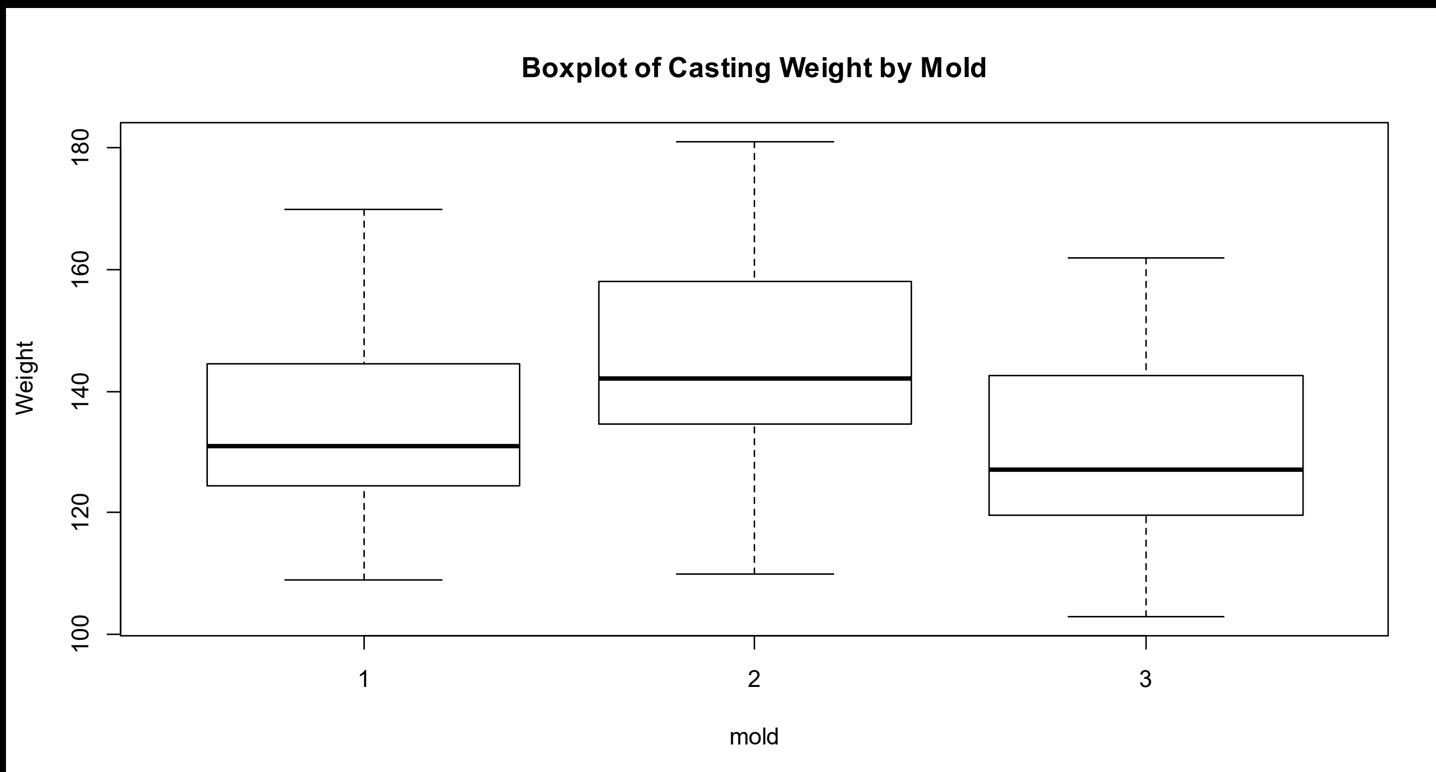
Boxplot to Compare Groups

```
> boxplot(y ~ x, data = data.frame)
```

Boxplot to Compare Groups

```
> boxplot(y ~ x, data = data.frame)
> boxplot(weight ~ mold, data = castings3)
```

Boxplot to Compare Groups



Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005