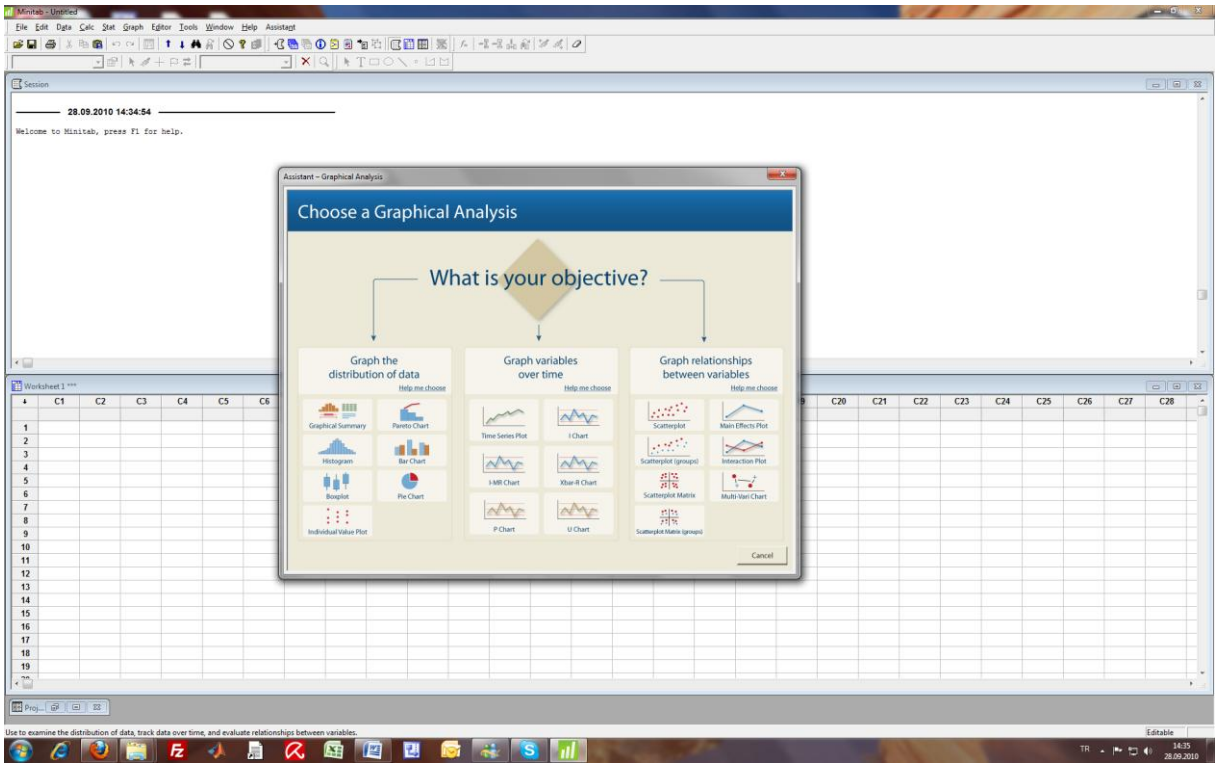
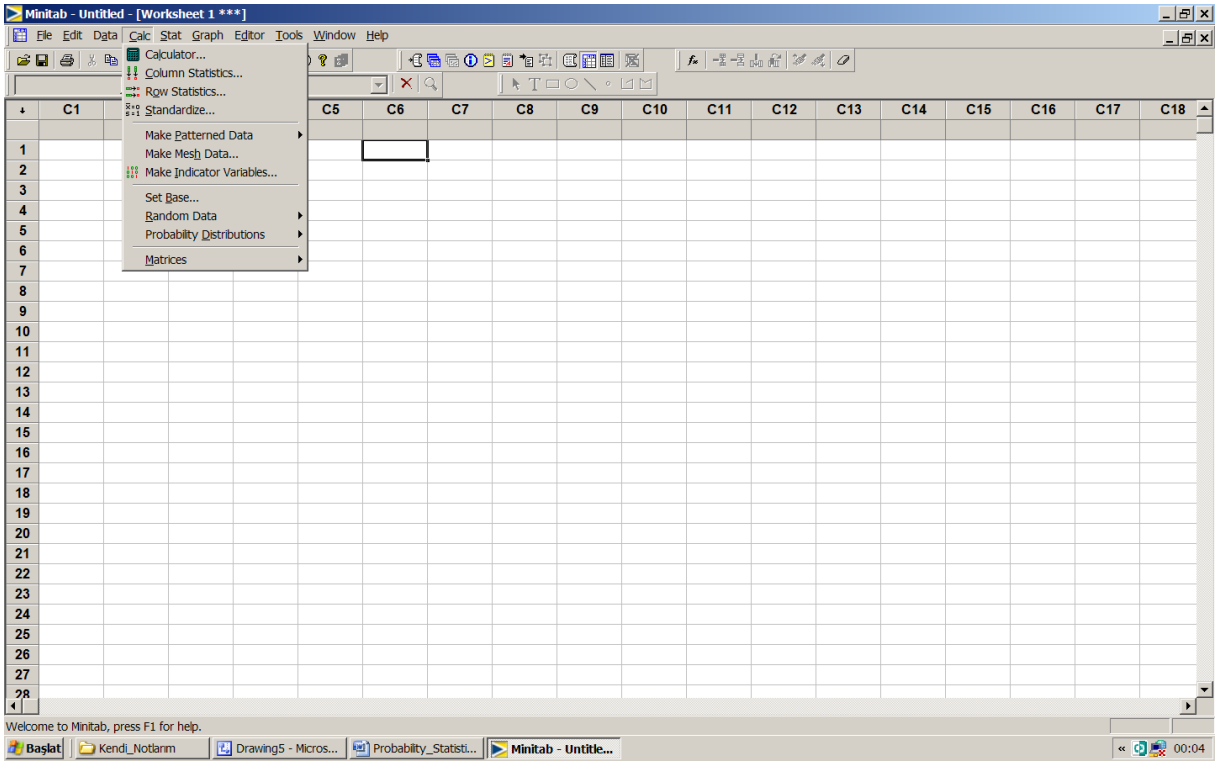
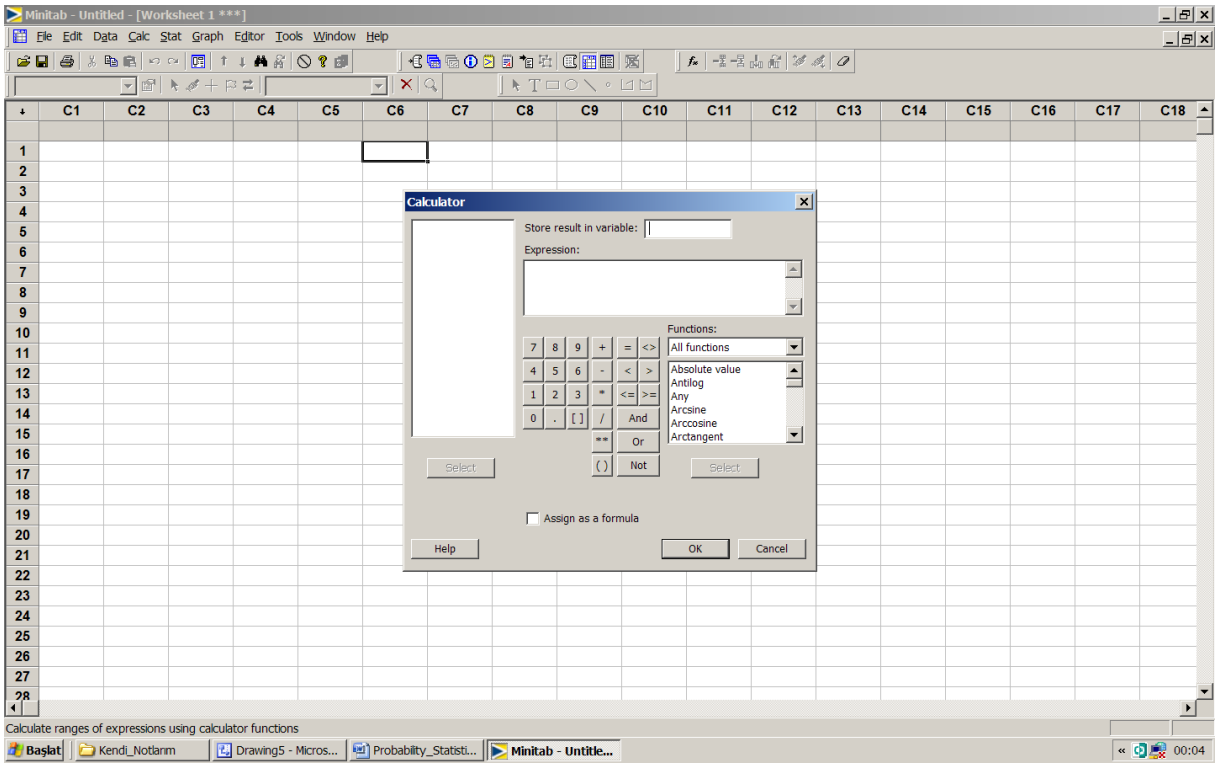
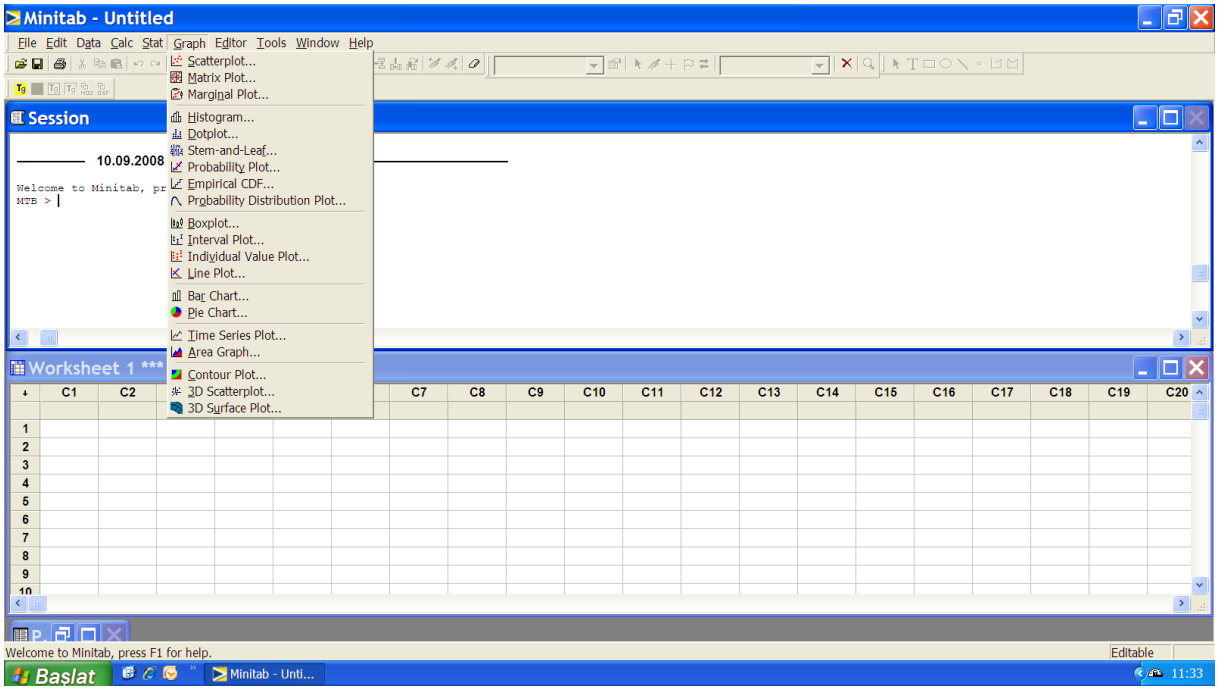
Graphs
For
Categorical Data

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes


Minitab - Untitled
28.09.2010 14:34:54
Welcome to Minitab, press F1 for help.
Assistant - Graphical Analysis
Choose a Graphical Analysis
What is your objective?
Graph the distribution of data
Graph variables over time
Graph relationships between variables
Cancel


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help
Calculator...
Column Statistics...
Row Statistics...
Standardize...
Make Patterned Data
Make Mesh Data...
Make Indicator Variables...
Set Base...
Random Data
Probability Distributions
Matrices
Welcome to Minitab, press F1 for help.
Başlat
Kendi_Notlarım
Drawing5 - Micros...
Probability_Statisti...
Minitab - Untitle...
00:04


Minitab - Untitled - [Worksheet 1 ***]
File Edit Data Calc Stat Graph Editor Tools Window Help
Calculator
Store result in variable:
Expression:
Functions:
All functions
Absolute value
Antilog
Any
Arcsine
Arccosine
Arctangent
Select
Select
Assign as a formula
Help
OK
Cancel
Calculate ranges of expressions using calculator functions
Başlat
Kendi_Notlarım
Drawing5 - Micros...
Probability_Statisti...
Minitab - Untitle...
00:04


Minitab - Untitled
File Edit Data Calc Stat Graph Editor Tools Window Help
Scatterplot...
Matrix Plot...
Marginal Plot...
Histogram...
Dotplot...
Stem-and-Leaf...
Probability Plot...
Empirical CDF...
Probability Distribution Plot...
Boxplot...
Interval Plot...
Individual Value Plot...
Line Plot...
Bar Chart...
Pie Chart...
Time Series Plot...
Area Graph...
Contour Plot...
3D Scatterplot...
3D Surface Plot...
Session
10.09.2008
Welcome to Minitab, pr
MTB >
Worksheet 1 ***
Welcome to Minitab, press F1 for help.
Editable
Başlat
Minitab - Unti...
11:33

After the data have been collected, they can be consolidated and summarized to show the following information:

- **What values of the variable have been measured?**
- **How often each value has occurred?**

For this purpose, we can construct a *statistical table* that can be used to display the data graphically as a data distribution. The type of graph we choose depends on the type of variable we have measured.

When the variable of interest is qualitative the statistical table is a list of the categories being considered along with a measure of how often each value occurred. We can measure "how often" in three different ways:

- **The frequency, or number of measurements in each category**
- **The relative frequency, or proportion of measurements in each category**
- **The percentage of measurements in each category**

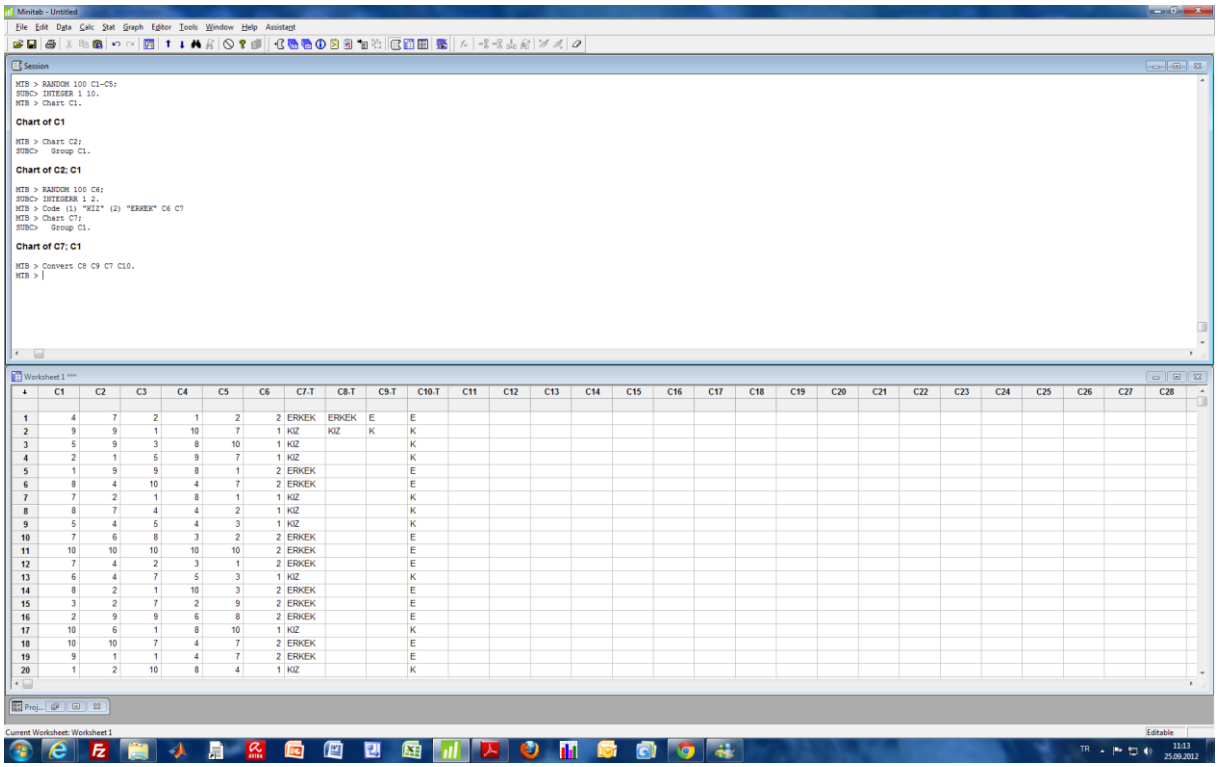
Frequency Distribution

Raw Data

Brown	Green	Brown	Blue	Yellow	Red	Blue	Red
Orange	Green	Blue	Brown	Black	Red	Blue	Green
Blue	Red	Blue	Red	Black	Black	Orange	Red
Red	Green	Red	Green				

Tally for Discrete Variables: Raw Data

Raw Data	Count	CumCnt	Percent	CumPct
Black	3	3	10.71	10.71
Blue	6	9	21.43	32.14
Brown	3	12	10.71	42.86
Green	5	17	17.86	60.71
Orange	2	19	7.14	67.86
Red	8	27	28.57	96.43
Yellow	1	28	3.57	100.00
N=	28			



MINITAB

CODE

CONVERT

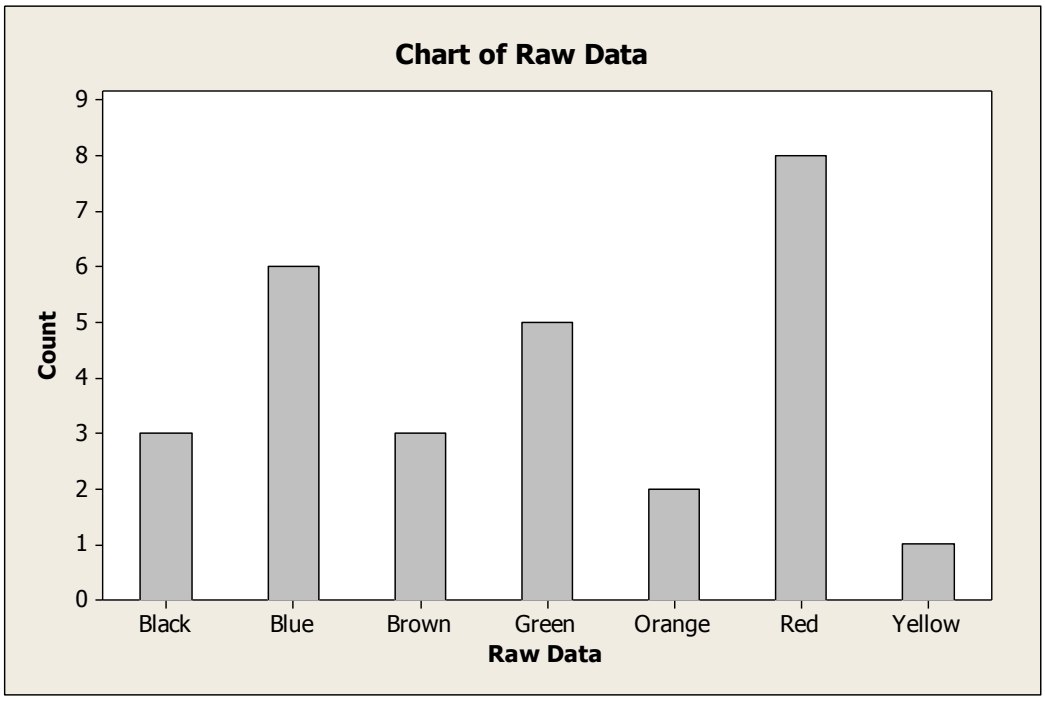
```
MTB > RANDOM 100 C6;
SUBC> INTEGERR 1 2.
MTB > Code (1) "KIZ" (2) "ERKEK" C6 C7
MTB > Chart C7;
SUBC>   Group C1.
```

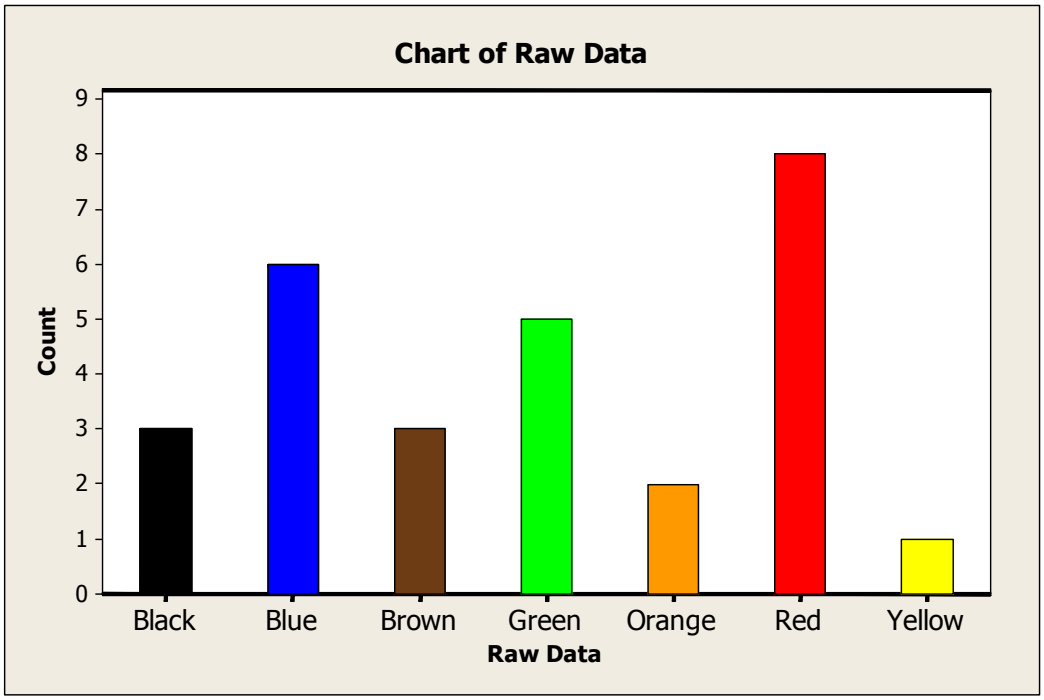
Chart of C7; C1

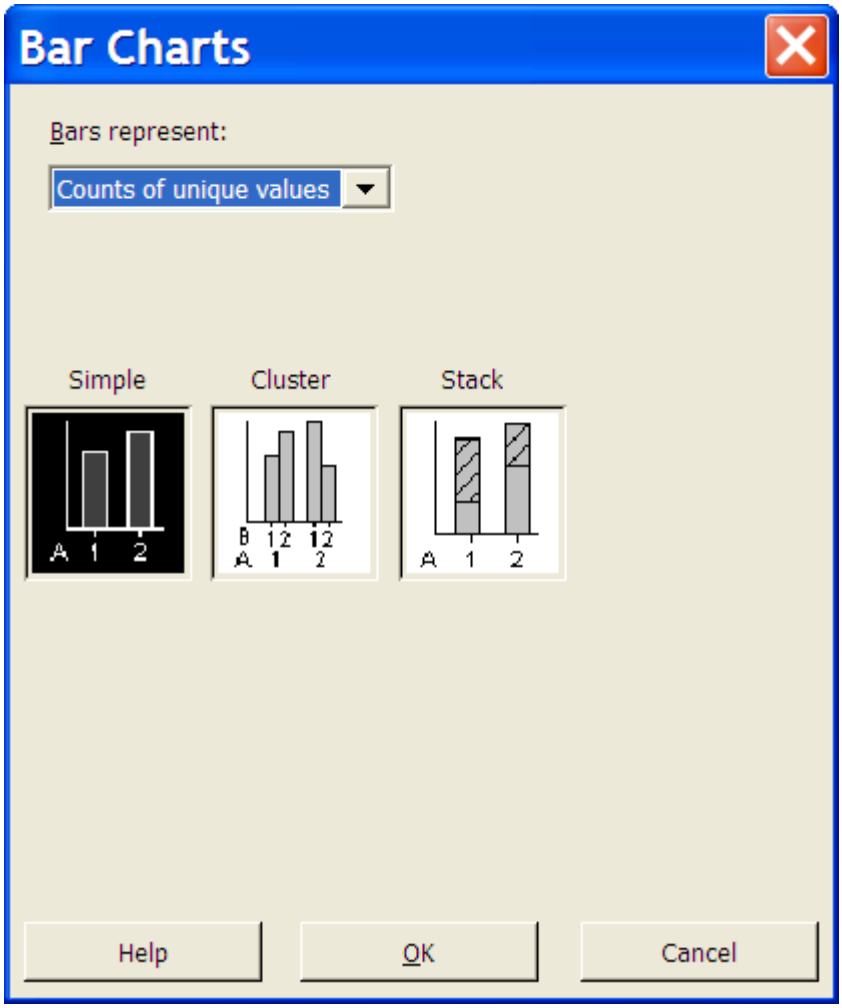
```
MTB > Convert C8 C9 C7 C10.
MTB >
```

Bar Chart

The bar chart is used to emphasize the actual quantity or frequency for each category.

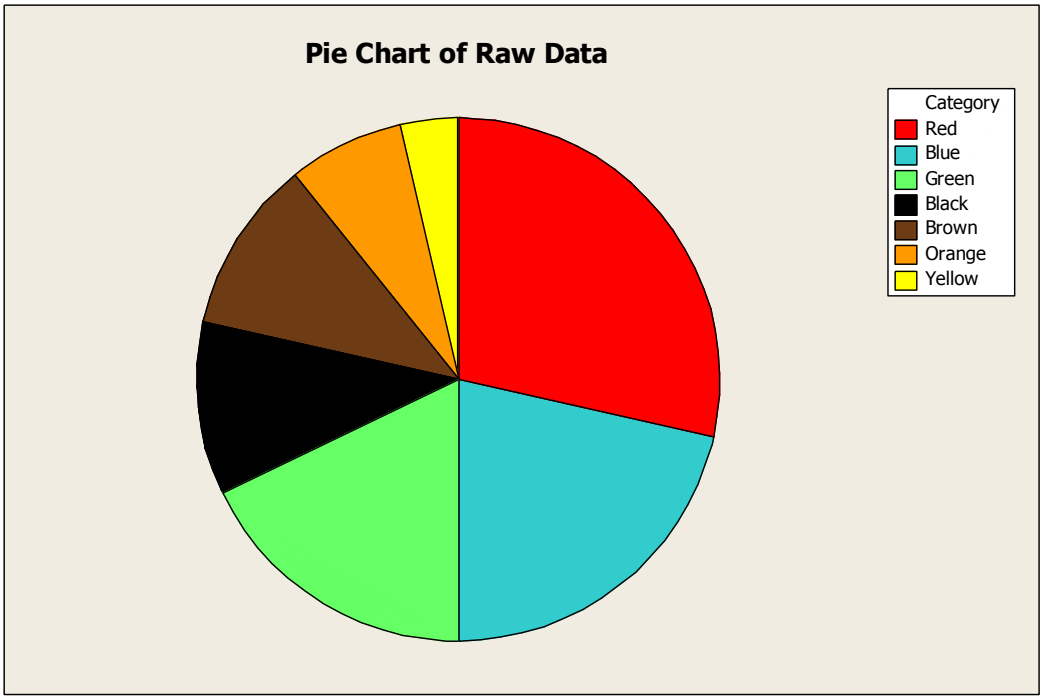


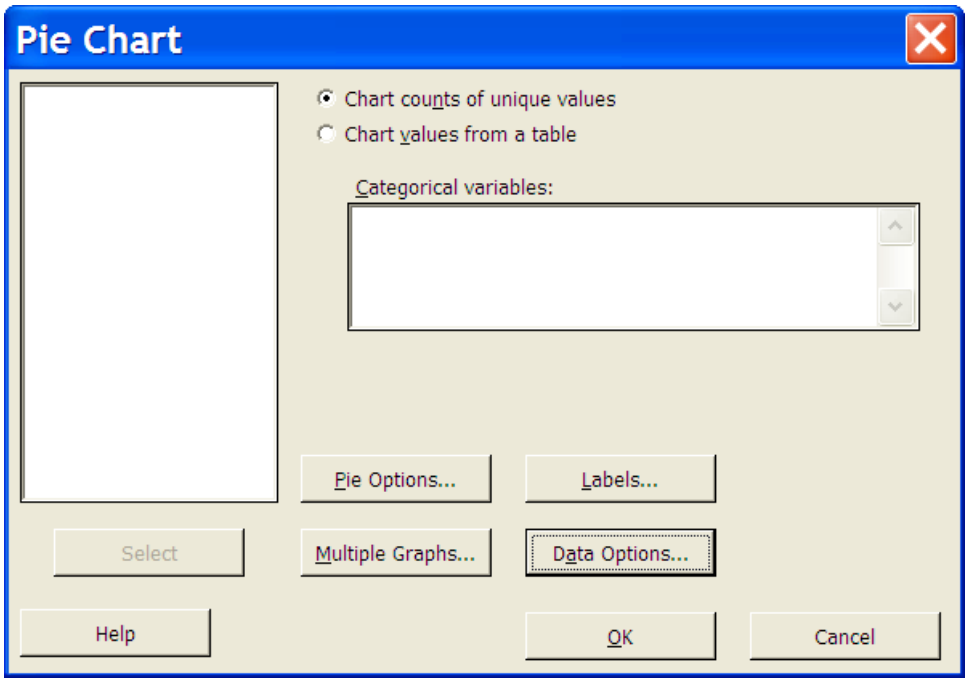




Pie Chart

The pie chat is used to display the relationship of the parts to the whole.

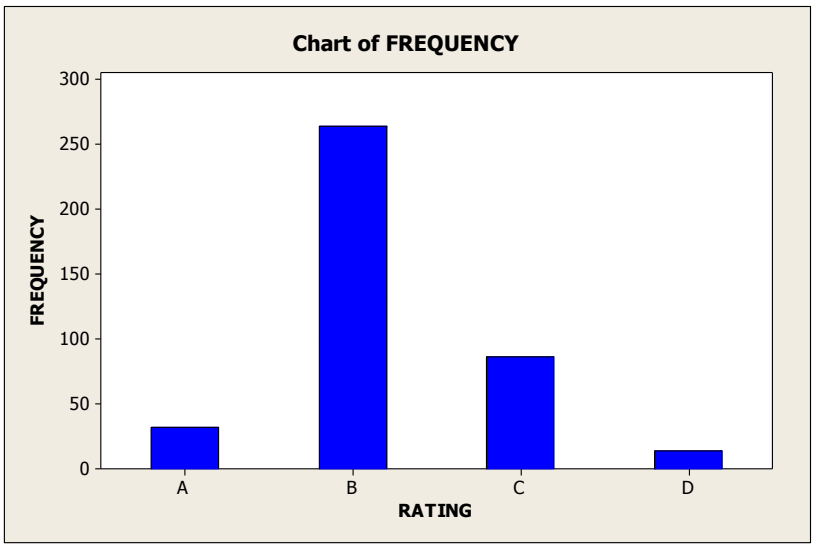


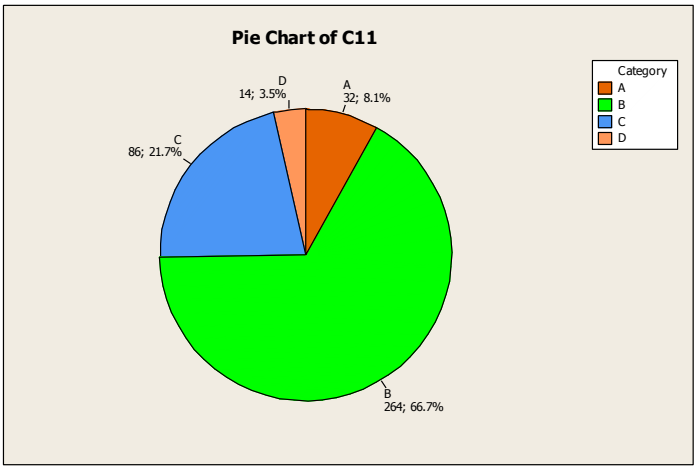


Values from a table

RATING	FREQUENCY
A	32

B	**264**
C	**86**
D	**14**
Total	**396**





Graphs
for Quantitative Data

Quantitative variables measure an amount or quantity on each experimental unit. If the variable can take only a finite or countable number of values, it is a discrete variable. A variable that can assume as infinite number of values corresponding to points on a line interval is called continuous.

Two Modes in MINITAB

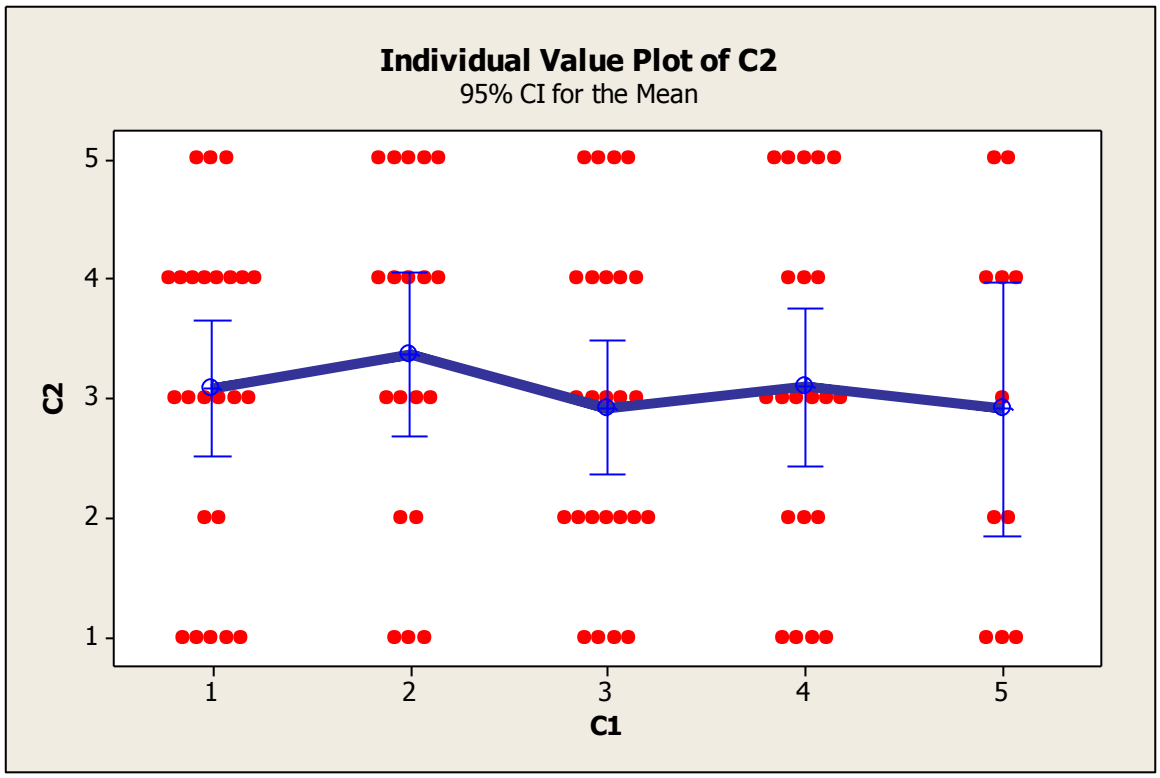
- GSTD (Standard Graphics)
- GPRO (Professional Graphics)

```
MTB > gstd

* NOTE * The character graph commands are obsolete.


* NOTE * Standard Graphics are now enabled, and Professional
Graphics are
        * disabled. Use the GPRO command when you want to re-
enable
        * Professional Graphics.

MTB > gpro

* NOTE * Professional Graphics are now enabled, and Standard
Graphics are
        * disabled. Use the GSTD command when you want to re-
enable Standard
        * Graphics.
MTB > table c1 c2
```

Tabulated statistics: C1; C2

Rows: C1 Columns: C2

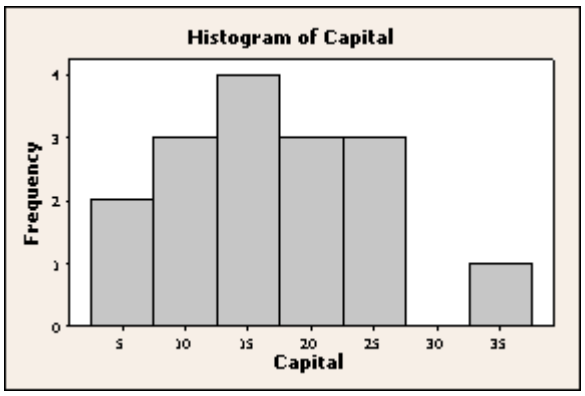
	1	2	3	4	5	All
1	5	2	6	8	3	24
2	3	2	4	5	5	19
3	4	7	5	5	4	25
4	4	3	6	3	5	21
5	3	2	1	3	2	11
All	19	16	22	24	19	100



Histogram

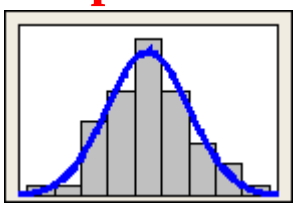
A graph used to assess the shape and spread of continuous sample data. You might create a histogram prior to or in conjunction with an analysis to help confirm assumptions and guide further analysis.

To draw a histogram Minitab divides sample values into many intervals called bins. By default, bars represent the number of observations falling within each bin (its frequency). In the histogram below, for example, there is one observation between one and two, four observations with values between two and three, and so on. Minitab automatically determines an optimal number of bins, but you can edit the number of bins as well as the intervals covered by each.

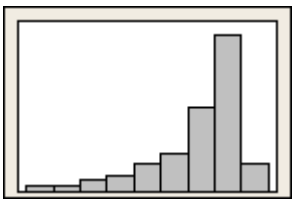


Here are some of the questions a histogram can help you answer:

Shape

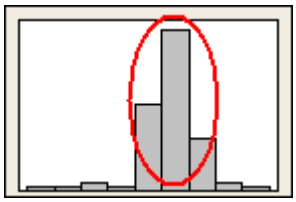


Do the data appear to be normally distributed?

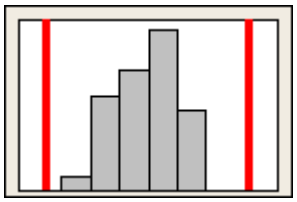


Do they skew left or right?

Spread



Are the data tightly clustered about a certain value?



Do the data stay within set limits?

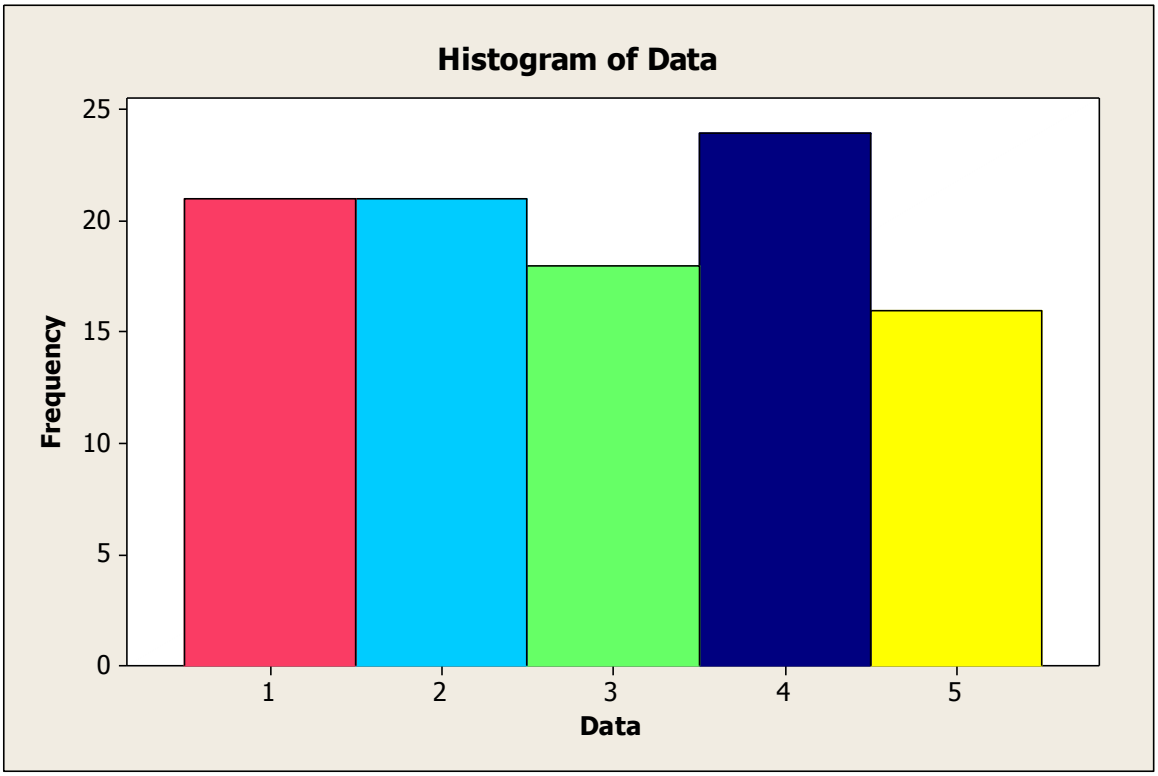
Data

2	5	2	2	5	2	5	3	5	4	2	1	3	4	4	4	1	3	1
3	5	2	3	2	4	4	1	2	4	4	4	1	3	2	1	5	3	4
4	5	3	5	2	1	2	1	1	3	1	5	4	1	4	2	1	1	4
4	2	2	5	2	5	4	3	5	3	2	3	5	4	5	3	3	2	3
2	3	1	4	1	4	2	5	4	4	3	1	4	1	3	4	1	1	4
1	2	5	1	2														

Tally for Discrete Variables: Data

Data	Count	CumCnt	Percent	CumPct
1	21	21	21.00	21.00
2	21	42	21.00	42.00
3	18	60	18.00	60.00
4	24	84	24.00	84.00
5	16	100	16.00	100.00
N=	100			

We may display a frequency distribution (or relative frequency distribution) graphically in the form of a histogram.



The Frequency Polygon

A frequency distribution can be portrayed graphically in yet another way by means of frequency polygon. To draw a frequency polygon we first place a dot above the midpoint of each class interval represented on the horizontal axis.

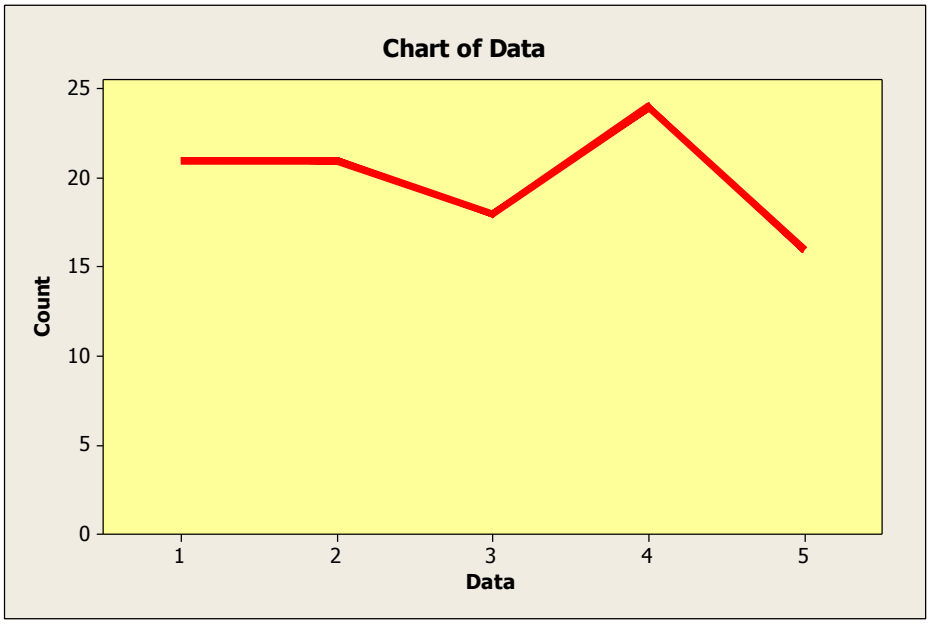
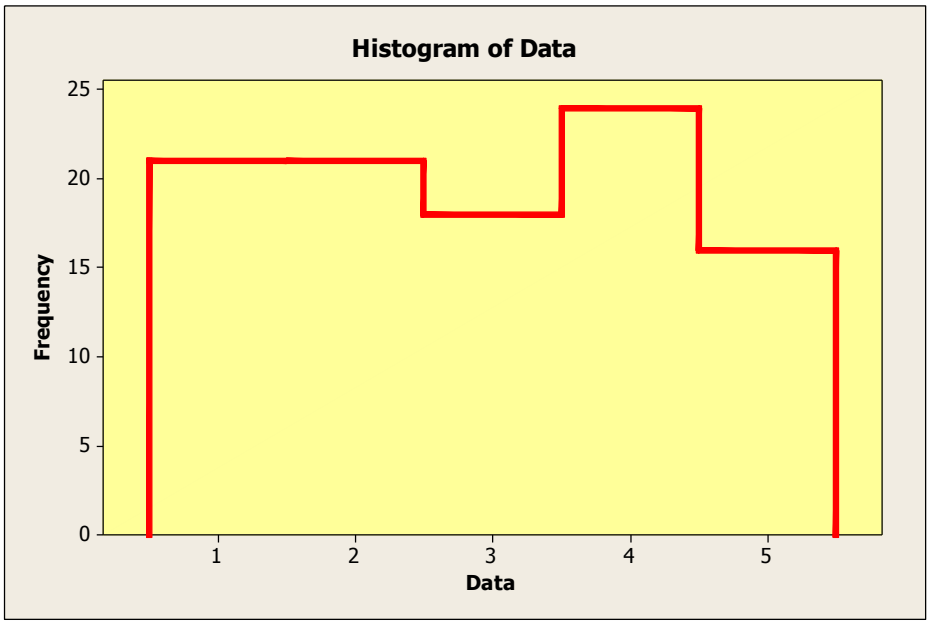


Chart of Data
Count
25
20
15
10
5
0
1
2
3
4
5
Data


Histogram of Data
Frequency
25
20
15
10
5
0
1
2
3
4
5
Data

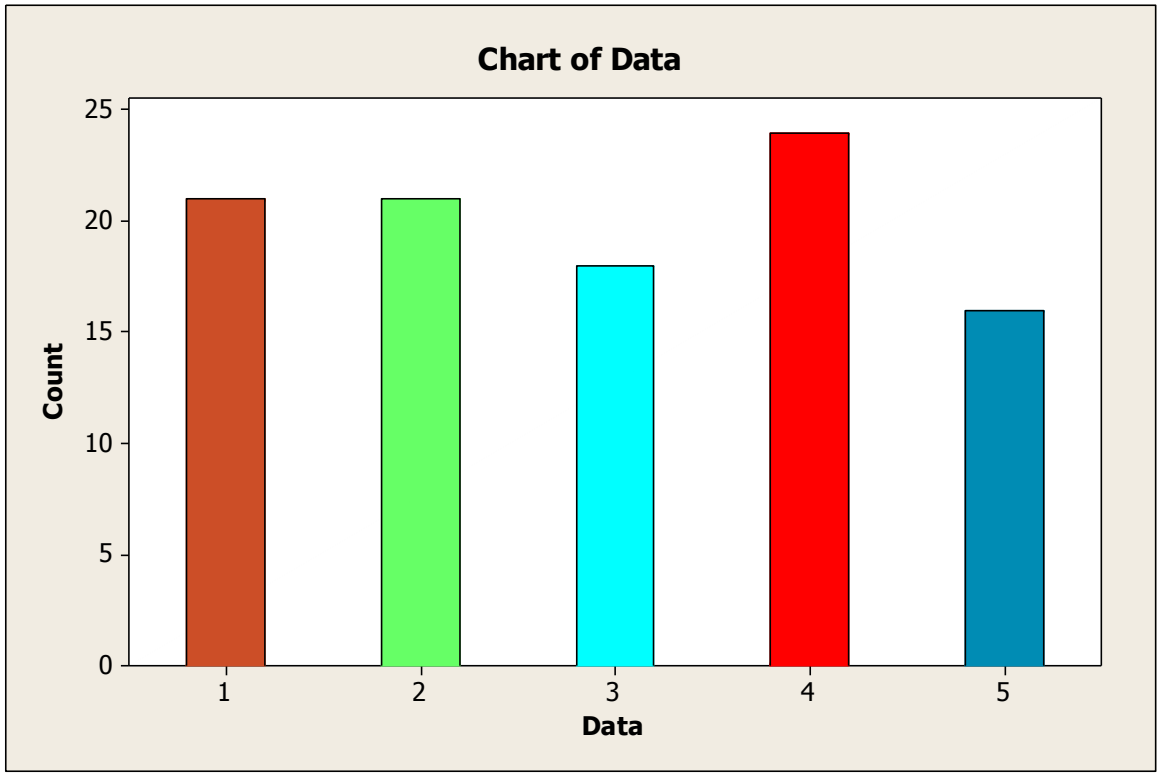
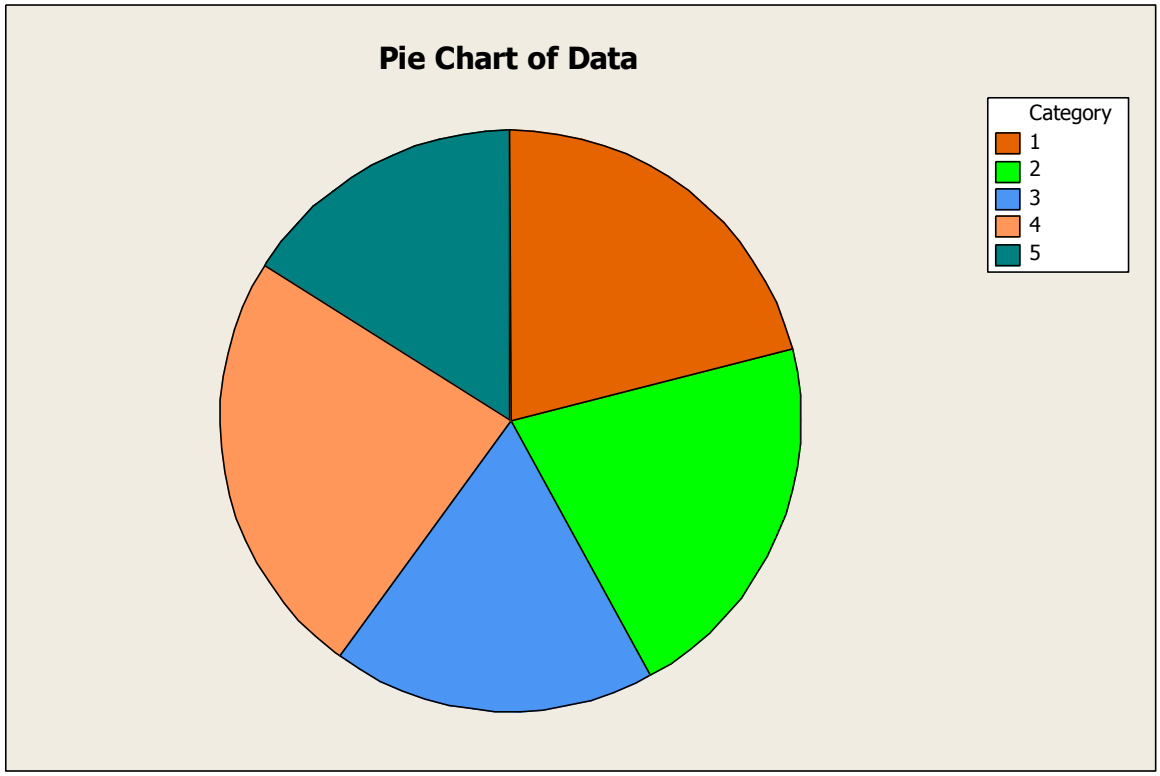


Chart of Data
Count
25
20
15
10
5
0
1
2
3
4
5
Data


Pie Chart of Data
Category
1
2
3
4
5

Data N = 100

149.945	154.835	156.545	144.071	163.026	162.135	165.427
148.730	149.262	164.053	164.495	158.067	168.213	154.498
153.786	179.842	156.820	155.380	165.987	165.438	161.087
149.182	157.426	156.463	178.250	144.227	155.331	159.765
138.103	172.821	148.118	132.886	153.671	143.240	168.314
154.841	154.002	165.417	181.485	166.551	171.861	162.939
155.172	170.734	142.544	152.615	174.651	145.293	159.152
147.413	163.648	148.095	168.324	151.095	167.299	165.087
152.384	160.981	157.146	174.677	160.379	163.586	153.406
170.266	159.349	167.332	174.977	152.617	170.006	171.099
162.280	157.812	179.336	169.194	177.780	162.365	159.144
155.067	159.010	163.244	146.199	151.624	162.675	145.991
158.975	140.753	157.112	179.546	148.888	156.763	155.333
165.635	164.900	172.124	166.485	154.711	159.680	155.335
168.676	167.275					

Histogram of Data

Midpoint Count

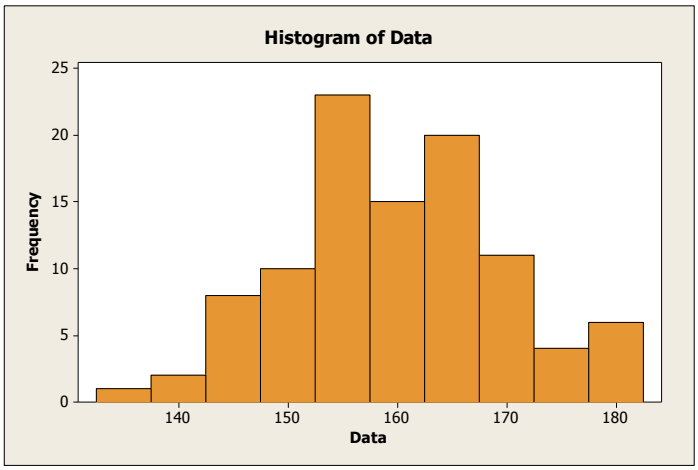
135	1	*
140	2	**
145	8	********
150	10	**********
155	23	***********************
160	15	***************
165	20	********************
170	11	**********
175	4	****
180	6	******

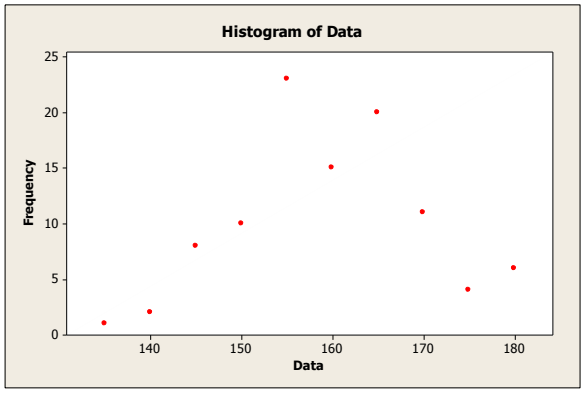
With different increment size

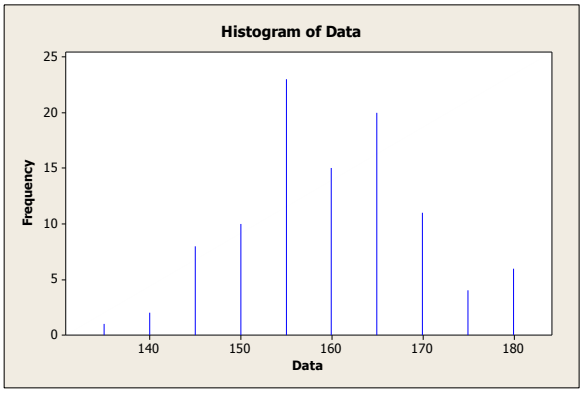
SUBC> increment 10.

Midpoint Count

130.0	1	*
140.0	6	******
150.0	24	************************
160.0	37	************************************
170.0	26	**************************
180.0	6	******

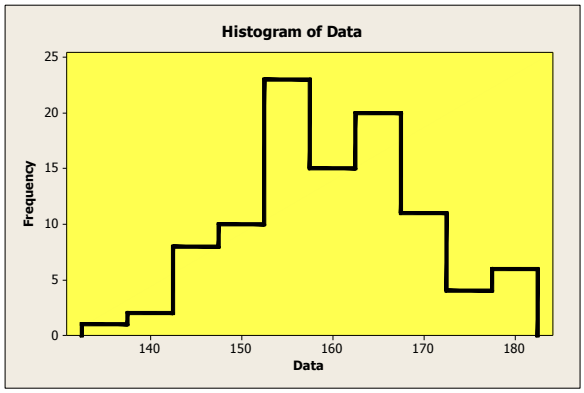


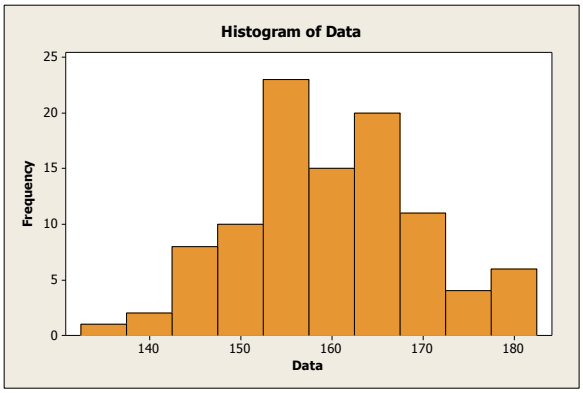




With symbols

With lines

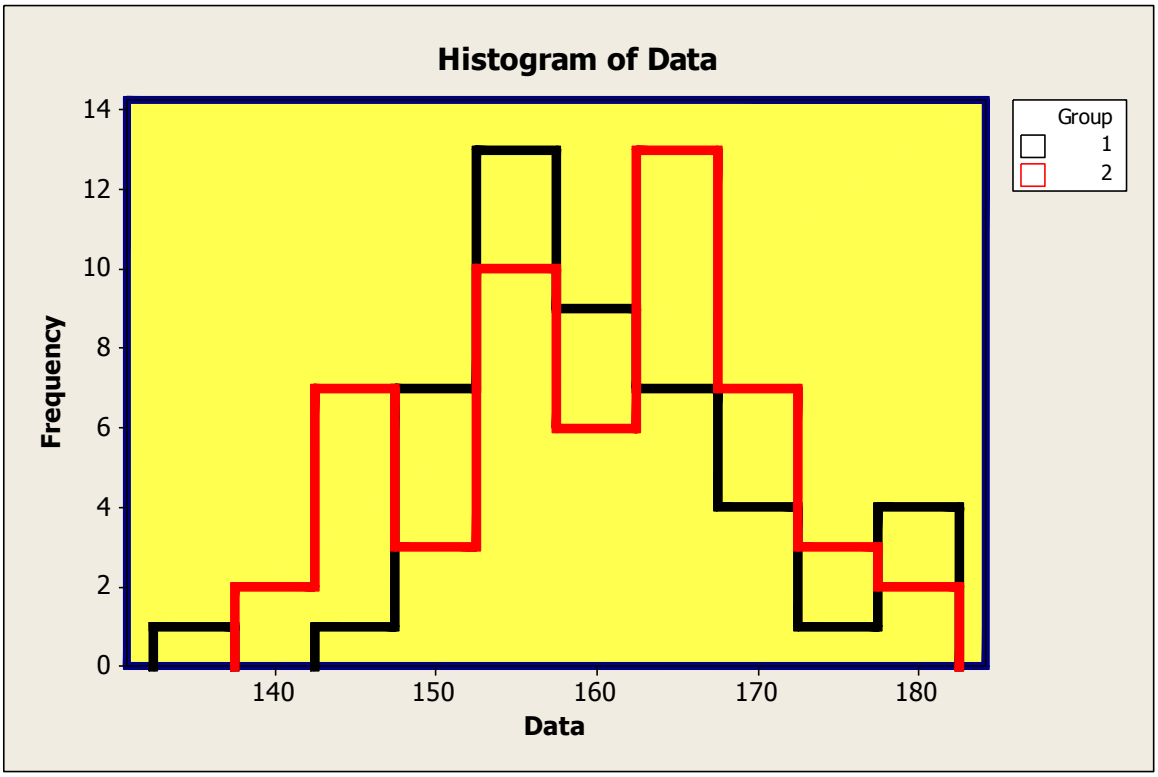




With area

With Groups

Data	Group
149.945	1
154.835	1
156.545	1
144.071	2
163.026	1
162.135	2
165.427	1



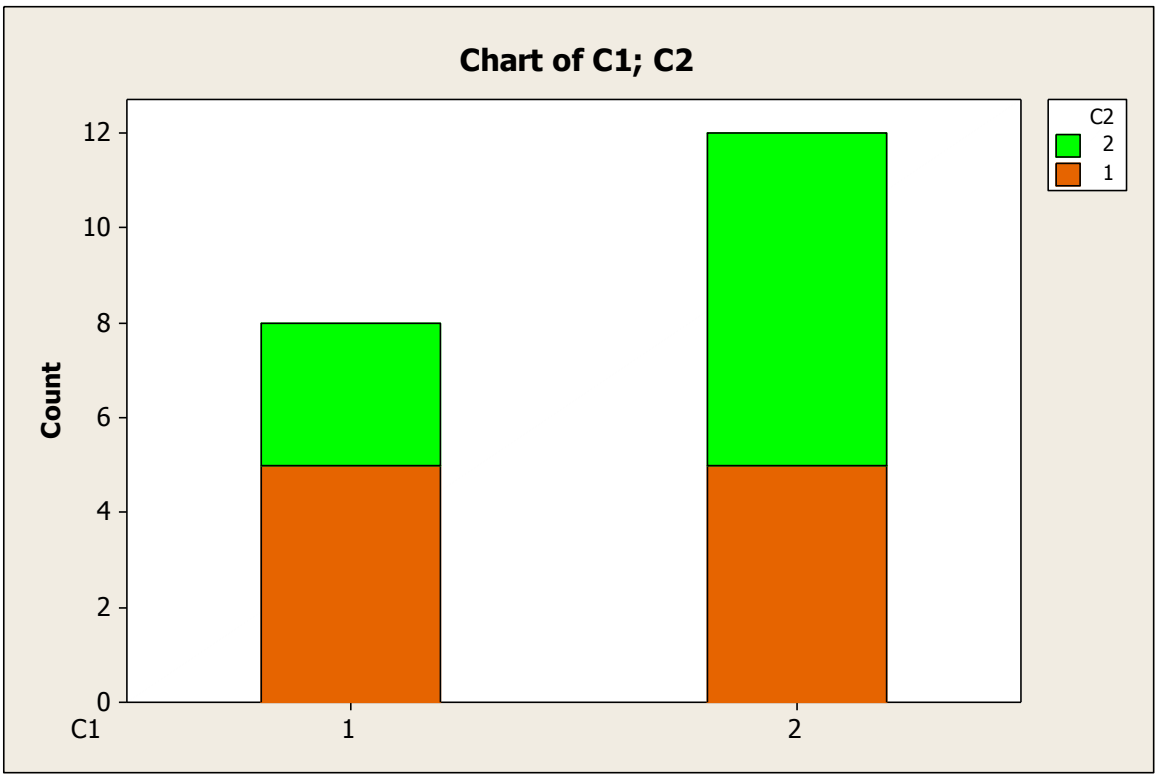
Histogram of Data Group = 1 N = 47

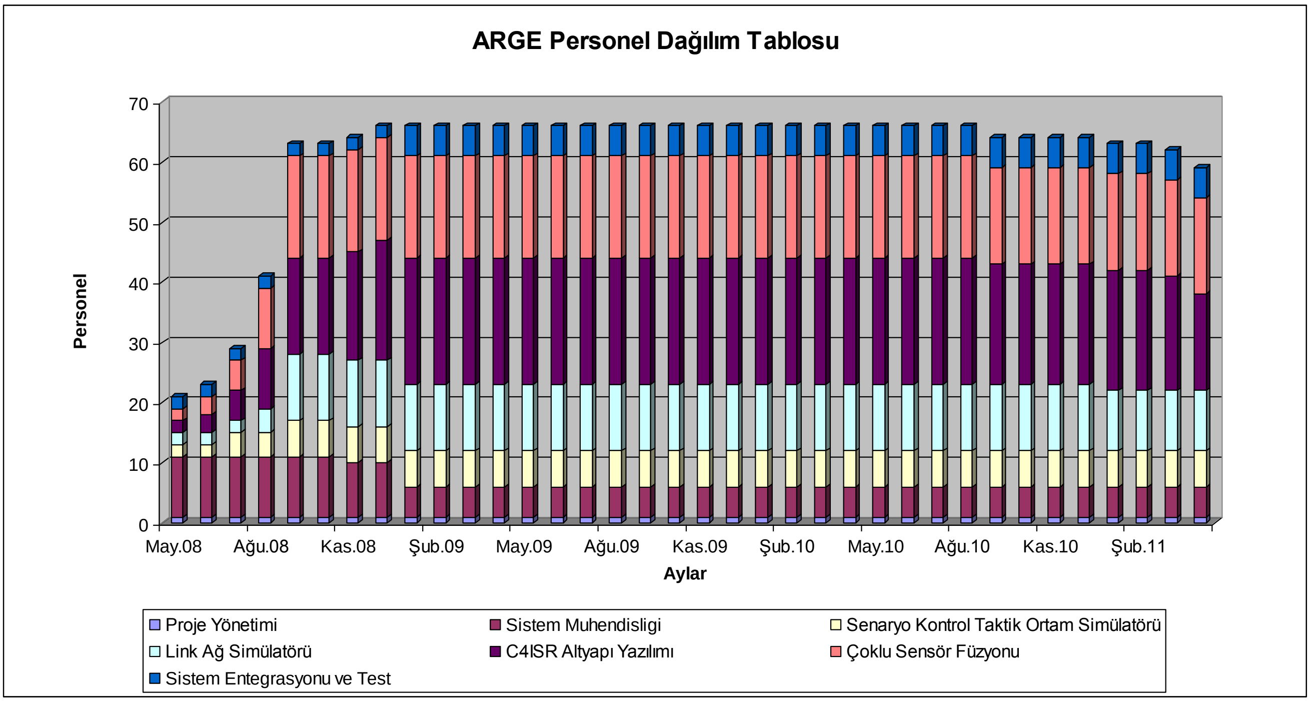
Midpoint	Count	
135	1	*
140	0	
145	1	*
150	7	*******
155	13	**********
160	9	*********
165	7	*******
170	4	****
175	1	*
180	4	****

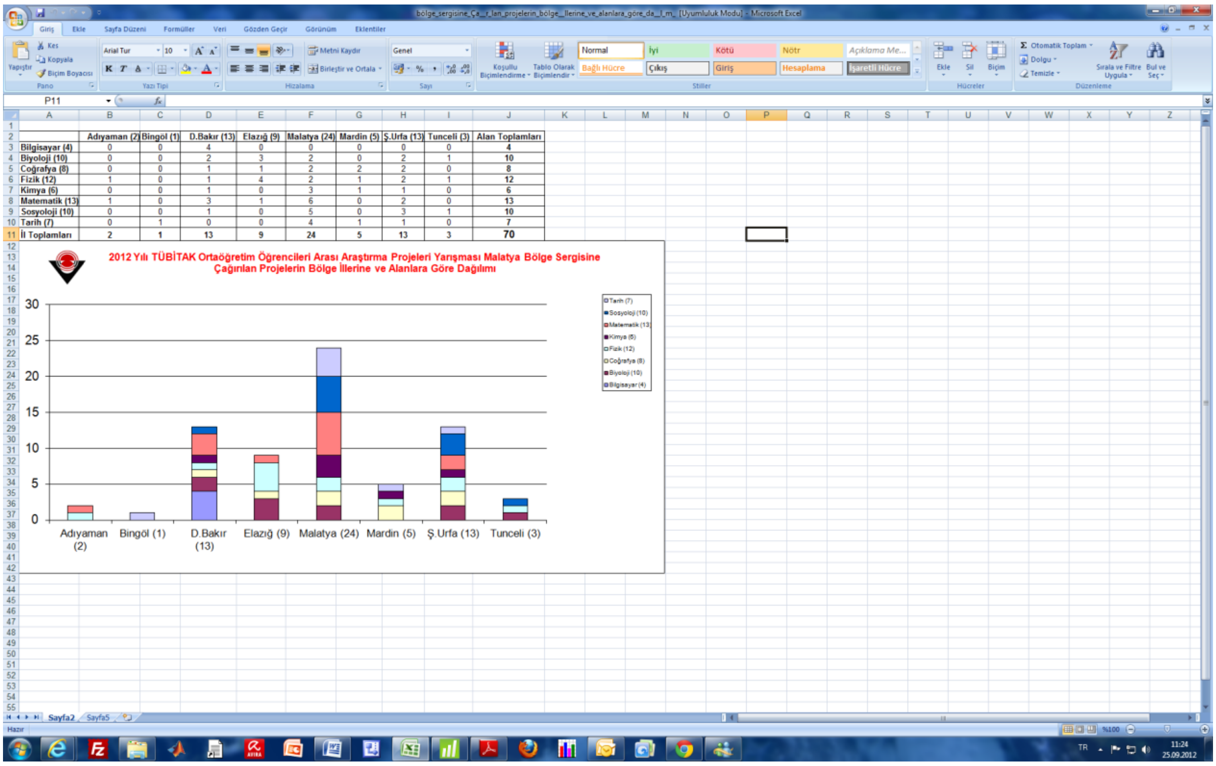
Histogram of Data Group = 2 N = 53

Midpoint	Count	
135	0	
140	2	**
145	7	*******
150	3	***
155	10	**********
160	6	******
165	13	**********
170	7	*******
175	3	***
180	2	**

```
Rows: C1    Columns: C2
          1    2  All
1         5    3    8
2         5    7   12
All      10   10   20
```







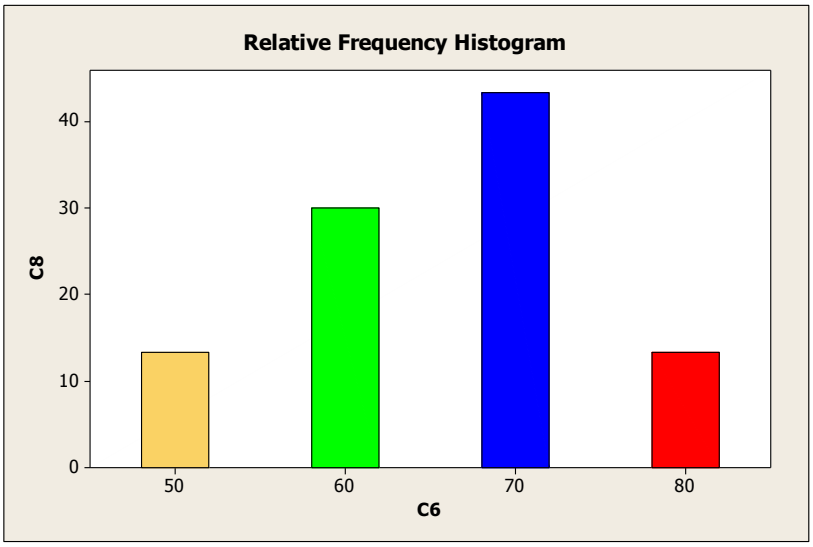
	Adıyaman (2)	Bingöl (1)	D.Bakır (13)	Elazığ (9)	Malatya (24)	Mardin (5)	Ş.Urfa (13)	Tunceli (3)	Alan Toplamları
Bilgisayar (4)	0	0	4	0	0	0	0	0	4
Biyoloji (10)	0	0	2	3	2	0	2	1	10
Coğrafya (8)	0	0	1	1	2	2	2	0	8
Fizik (12)	1	0	1	4	2	1	2	1	12
Kimya (6)	0	0	1	0	3	1	1	0	6
Matematik (13)	1	0	3	1	6	0	2	0	13
Sosyoloji (10)	0	0	1	0	5	0	3	1	10
Tarih (7)	0	1	0	0	4	1	1	0	7
İl Toplamları	2	1	13	9	24	5	13	3	70

Relative Frequency Histograms

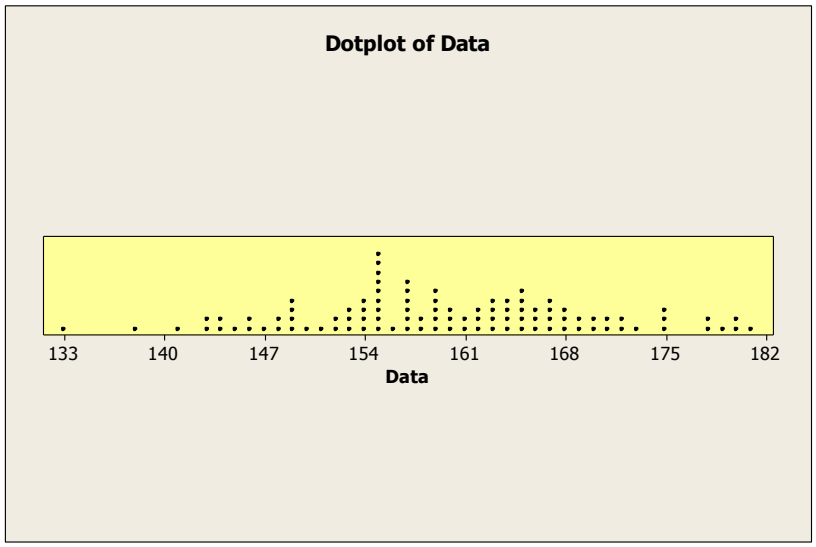
A relative frequency histogram for a quantitative data set is a bar graph in which the height of the bar shows "how often" (measured as a proportion or relative frequency) measurements fall in a particular class or subinterval.

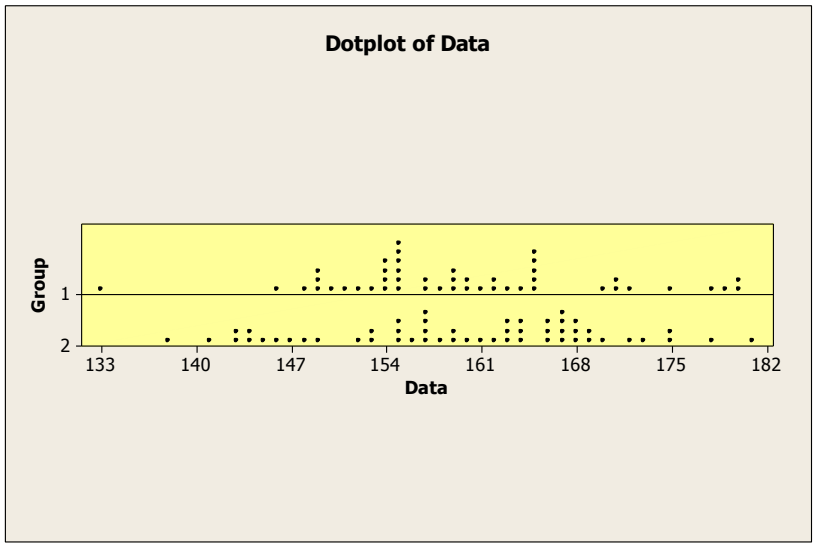
Data

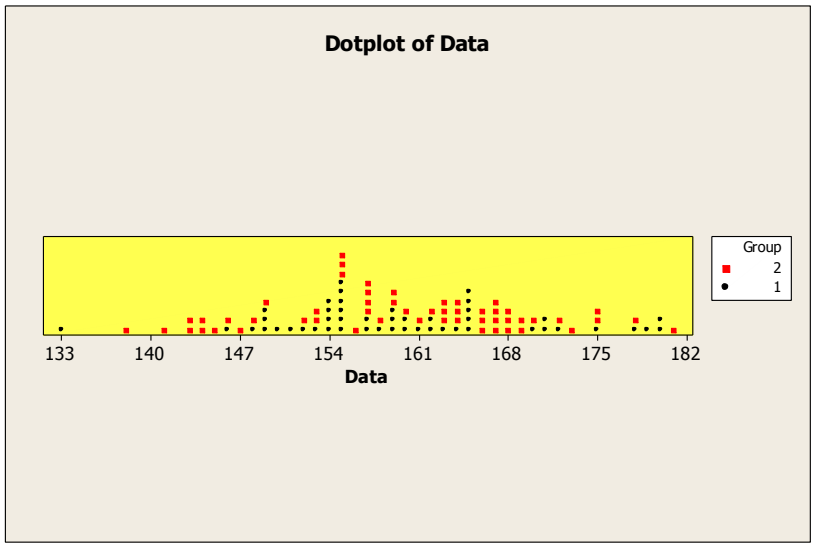
Row	interval	frequency	relative frequency
1	50	4	13.3333
2	60	9	30.0000
3	70	13	43.3333
4	80	4	13.3333



Dotplots


Dotplot of Data
133
140
147
154
161
168
175
182
Data


Dotplot of Data
Group
1
2
133
140
147
154
161
168
175
182
Data


Dotplot of Data
Group
2
1
133
140
147
154
161
168
175
182
Data

Stem and Leaf Plots

A simple way to display the distribution of a quantitative data set is the stem and leaf plot. This plot presents a graphical display of the data using the actual numerical values of each data point.

1. **Divide each measurement into two parts: the stem and leaf.**
2. **List the stems in a column, with a vertical line to their right.**
3. **For each measurement, record the leaf portion in the same row as its corresponding stem.**
4. **Order the leaves from lowest to highest in each stem.**
5. **Provide a key to your stem and leaf coding so that the reader can re-create the actual measurements if necessary.**

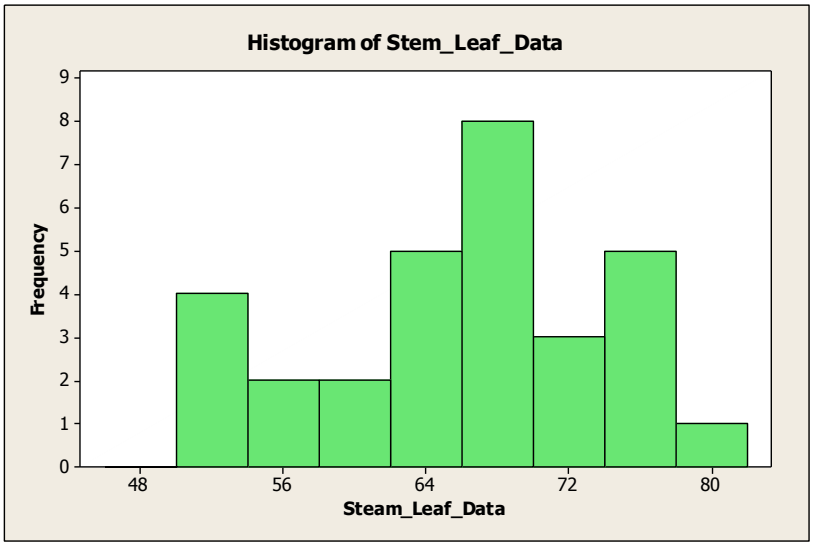
Stem_Leaf_Data

62	62	50	78	57	67	63	70	73	58	75	68	69	51	50
76	68	77	51	58	74	62	66	68	73	69	63	67	55	74

```
Stem-and-leaf of Steam_Leaf_Data  N  = 30
Leaf Unit = 1.0

 8     5  00115788
(13)   6  2223367788899
 9     7  033445678
```

Histogram

```
Histogram of Stem_Le   N = 30

Midpoint          Count
    50.0              4  ****
    60.0              9  *********
    70.0             13  *************
    80.0              4  ****
```

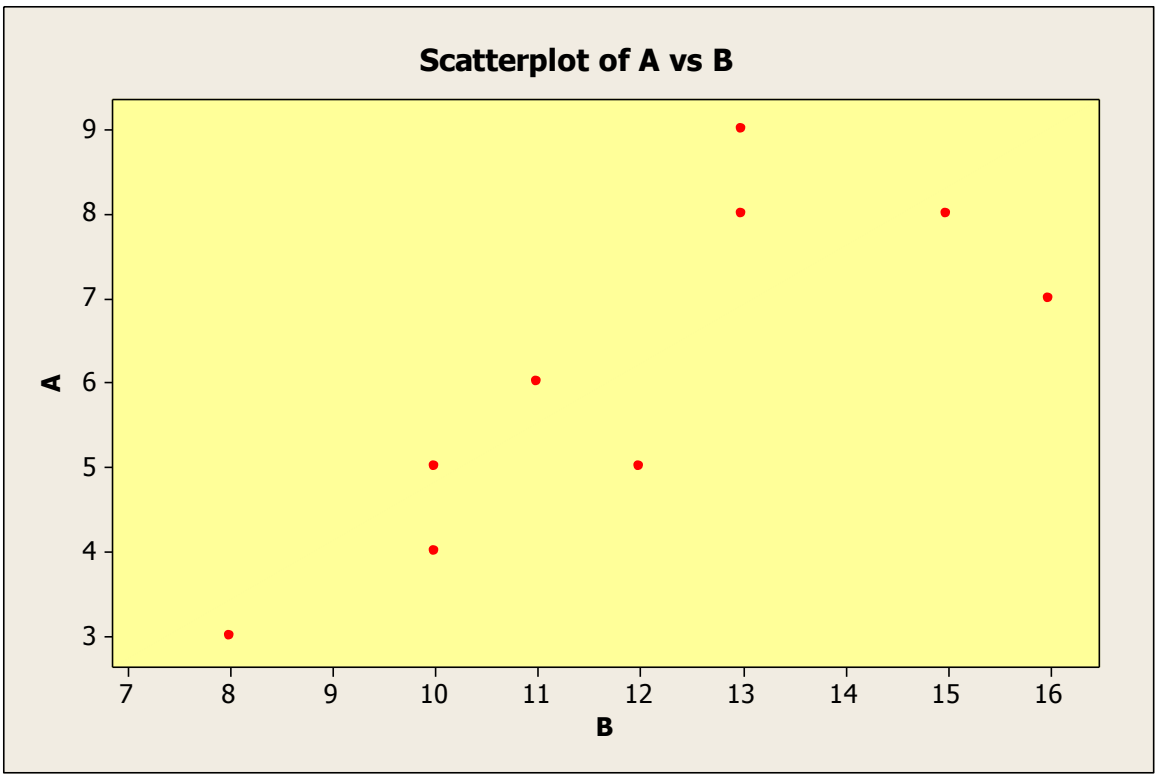


Data N = 100

149.945	154.835	156.545	144.071	163.026	162.135	165.427
148.730	149.262	164.053	164.495	158.067	168.213	154.498
153.786	179.842	156.820	155.380	165.987	165.438	161.087
149.182	157.426	156.463	178.250	144.227	155.331	159.765
138.103	172.821	148.118	132.886	153.671	143.240	168.314
154.841	154.002	165.417	181.485	166.551	171.861	162.939
155.172	170.734	142.544	152.615	174.651	145.293	159.152
147.413	163.648	148.095	168.324	151.095	167.299	165.087
152.384	160.981	157.146	174.677	160.379	163.586	153.406
170.266	159.349	167.332	174.977	152.617	170.006	171.099
162.280	157.812	179.336	169.194	177.780	162.365	159.144
155.067	159.010	163.244	146.199	151.624	162.675	145.991
158.975	140.753	157.112	179.546	148.888	156.763	155.333
165.635	164.900	172.124	166.485	154.711	159.680	155.335
168.676	167.275					

```
Stem-and-leaf of Data  N  = 100
Leaf Unit = 1.0
Increment 10


 2     13  28
 18    14  0234455678888999
(35)   15  11222333444445555556666777788999999
 47    16  00122222333344455555566777888 9
 16    17  000112244478999
 1     18  1
```

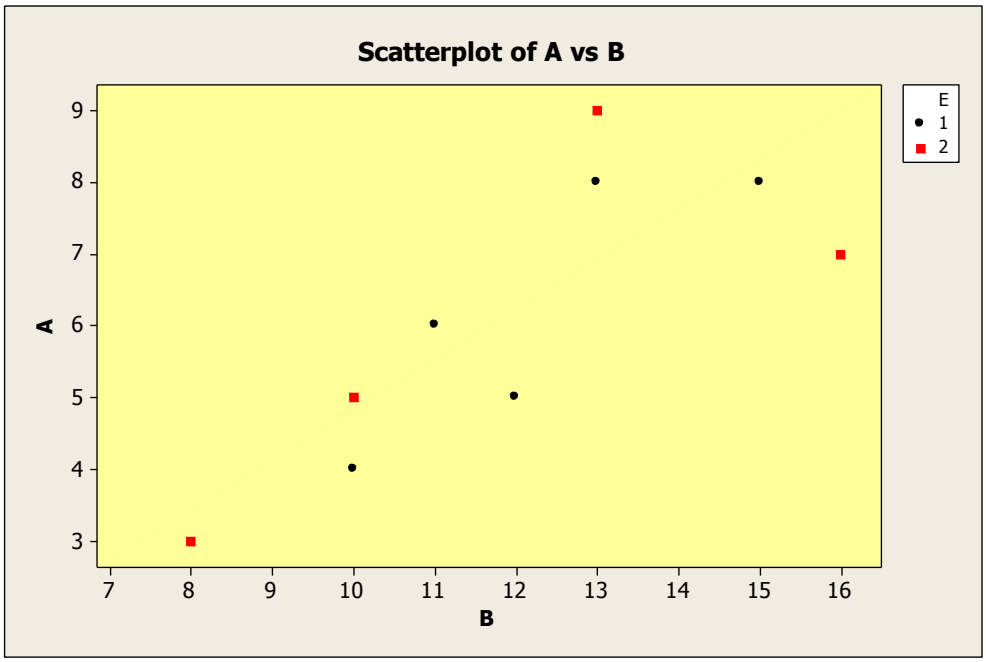
Scatter Plot

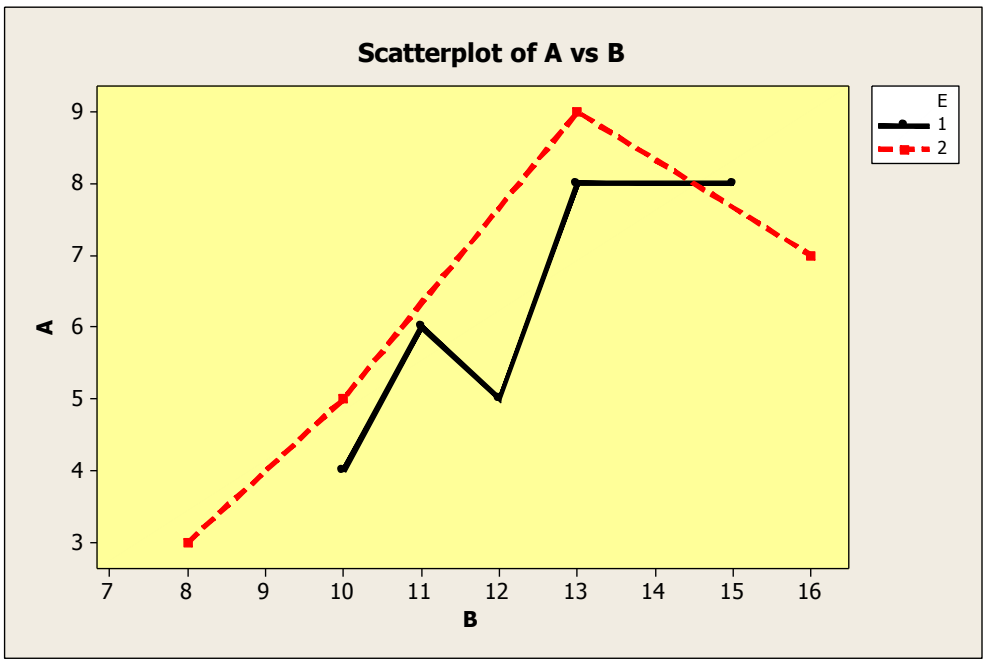


Data Display

Row	A	B
1	5	12
2	7	16
3	8	13
4	4	10
5	3	8
6	5	10
7	6	11
8	8	15
9	9	13

Scatter Plot with Groups



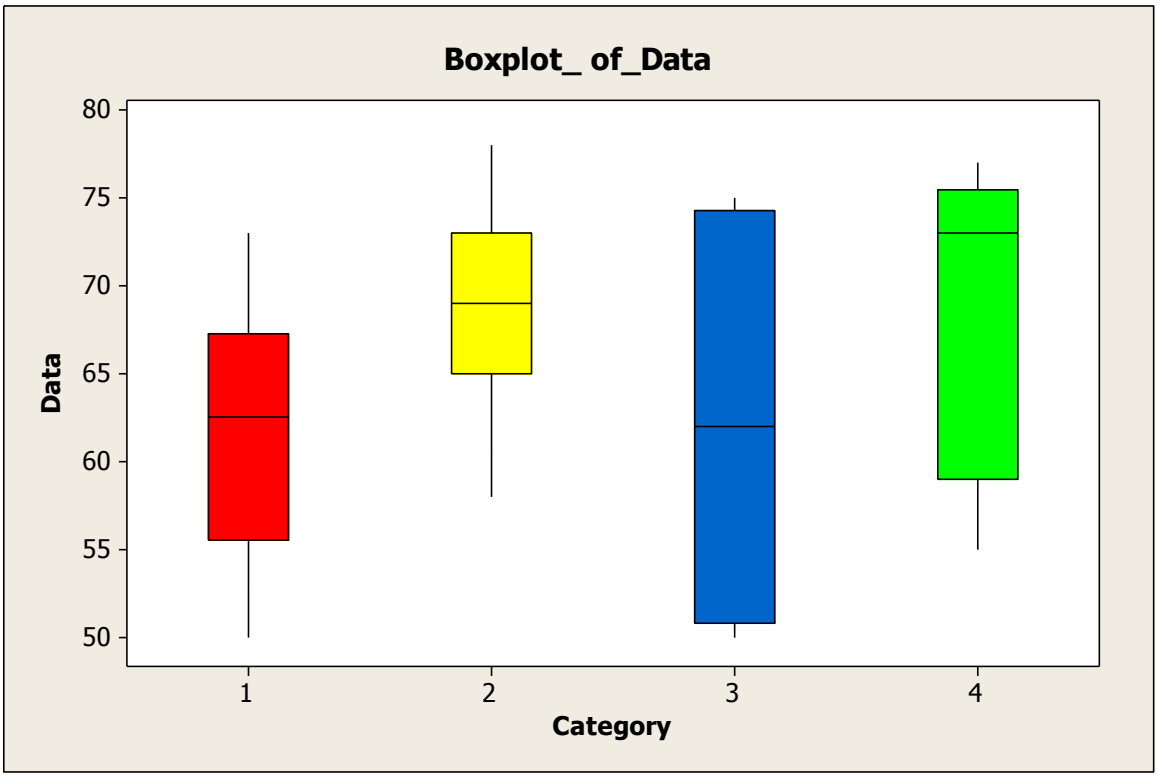


Data Display

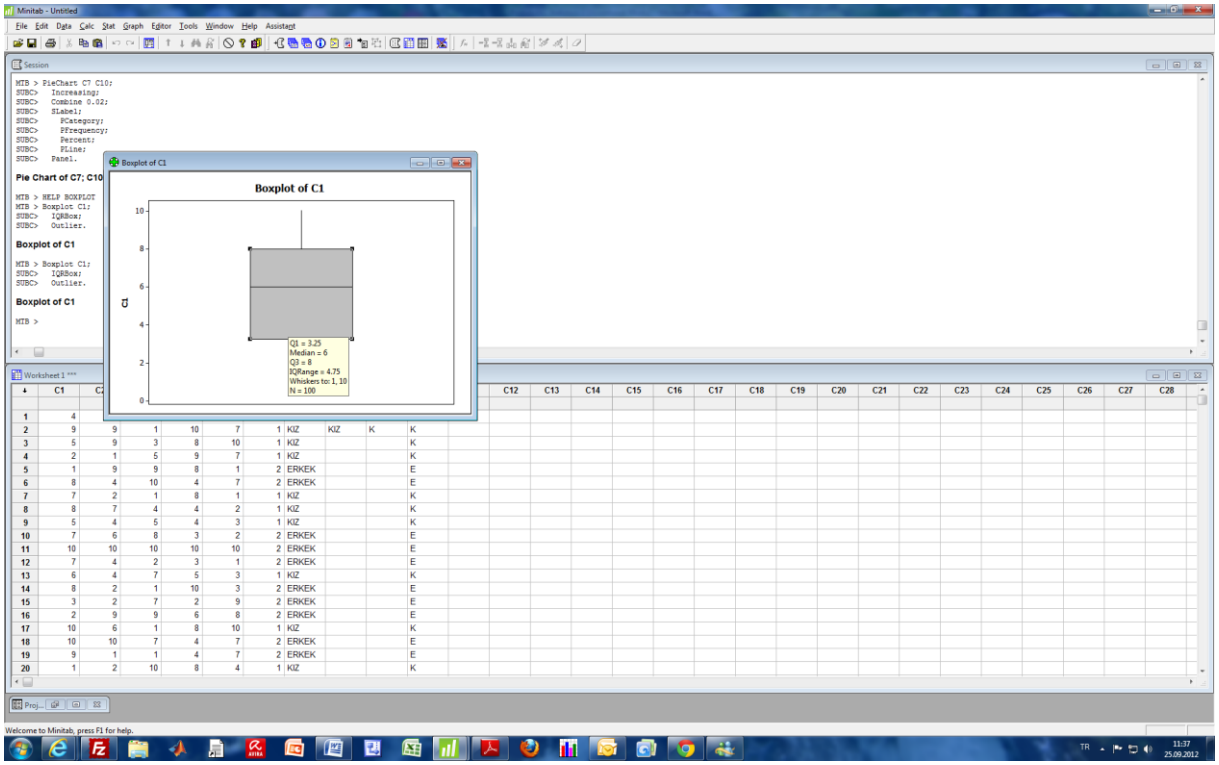
Row	A	B	E
1	5	12	1
2	7	16	2
3	8	13	1
4	4	10	1
5	3	8	2
6	5	10	2
7	6	11	1
8	8	15	1

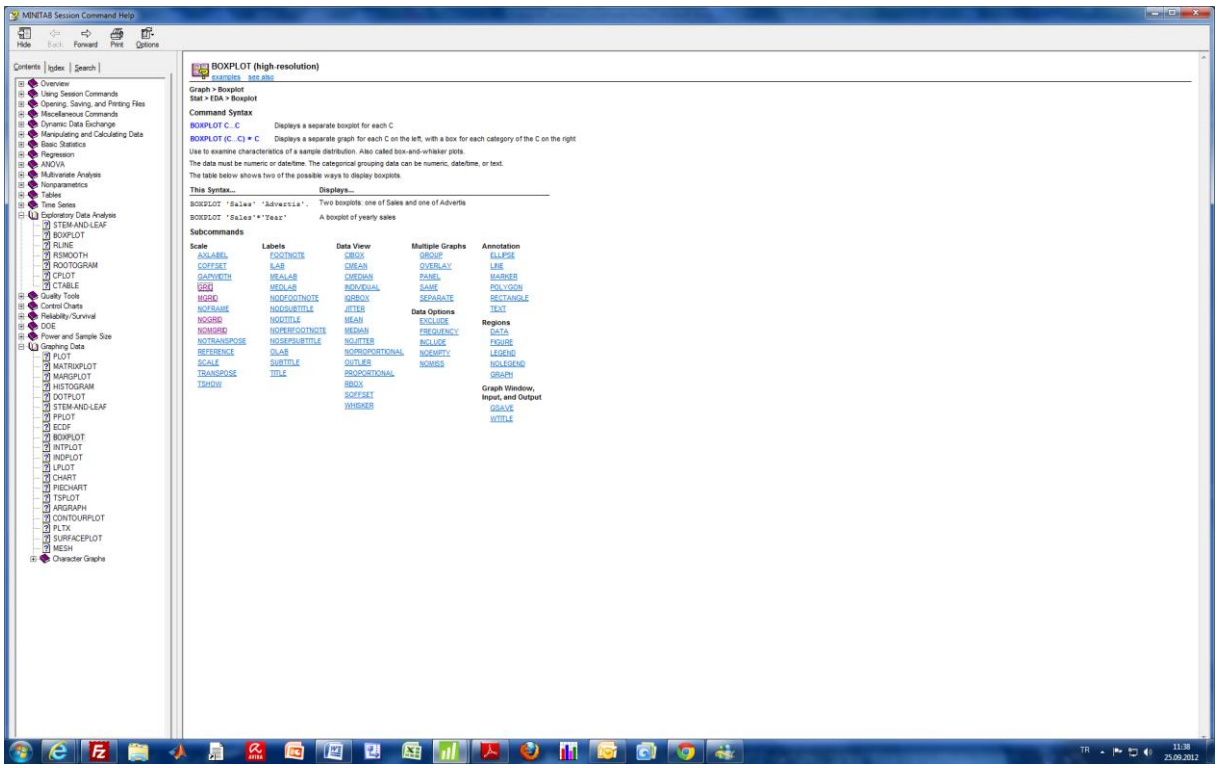
Box Plot

Use box plots (also called box-and-whisker plots) to assess and compare sample distributions. The figure below illustrates the components of a default box plot.



Row	Data	Category
1	62	1
2	62	2
3	50	3
4	78	2
.		
.		
.		
27	63	1
28	67	1
29	55	4
30	74	3


Minitab - Untitled
Boxplot of C1
Boxplot of C1
Q1 = 3.25
Median = 6
Q3 = 8
IQRange = 4.75
Whiskers to 1, 10
N = 100
Worksheet 1 ***
Welcome to Minitab, press F1 for help.


MINITAB Session Command Help
BOXPLOT (high-resolution)
Graph > Boxplot
Stat > EDA > Boxplot
Command Syntax
Subcommands

```
TALLY C4 C5
```

Tally for Discrete Variables: C4; C5

C4	Count	C5	Count
A	10	1	8
B	8	2	5
N=	18	3	5
		N=	18

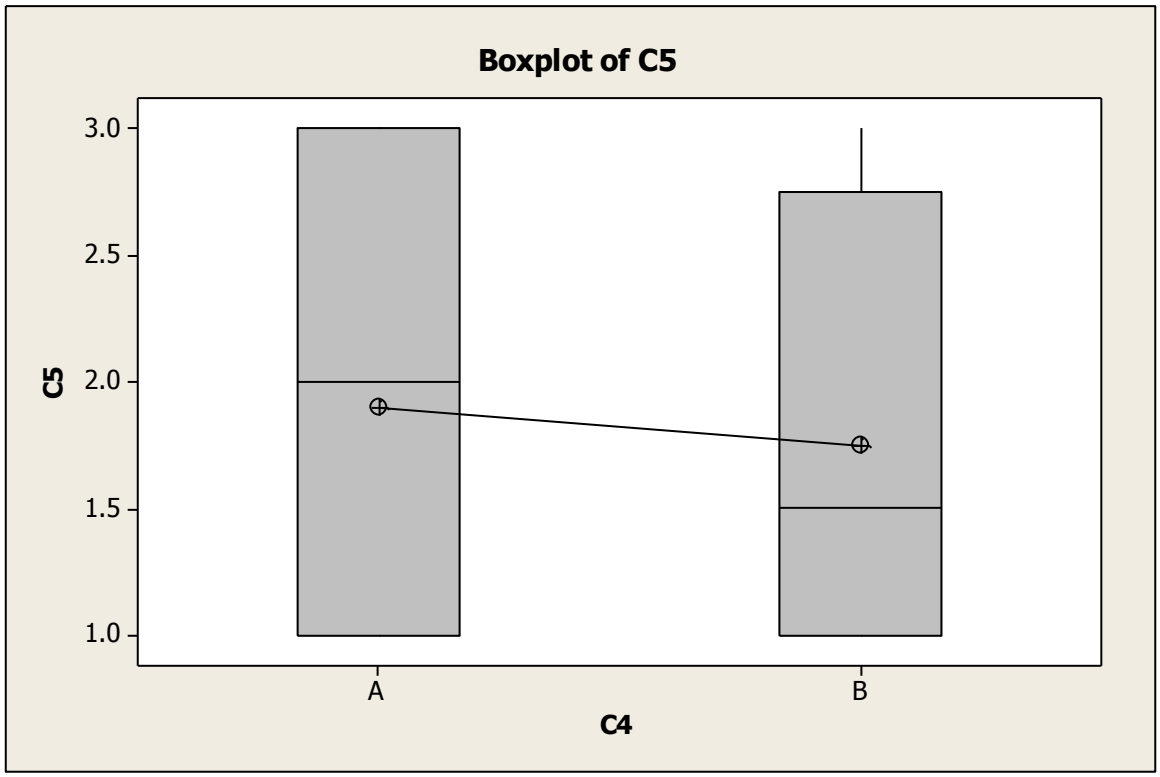
```
MTB > print c4
```

Data Display

```
C4
   A   A   A   B   B   A   A   B   B   A   B   B   A   A   A   A   B   B
MTB > print c5
```

Data Display

```
C5
   1   2   1   2   3   1   2   3   2   3   1   1   1   2   3   3   1   1
```

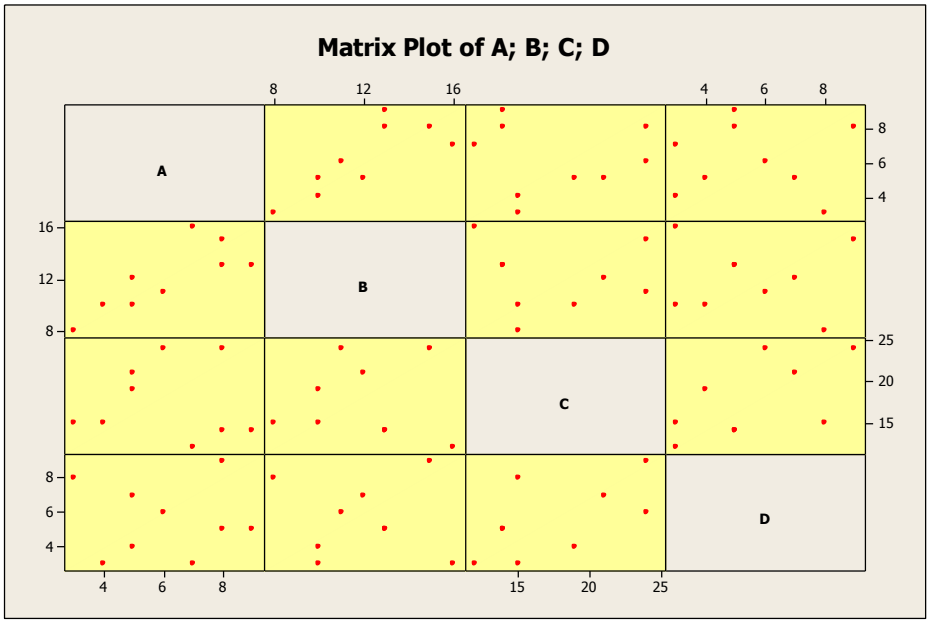


Matrix Plot

Assess the relationships between many pairs of variables at once by creating an array of scatter plots. There are two types of matrix plots:

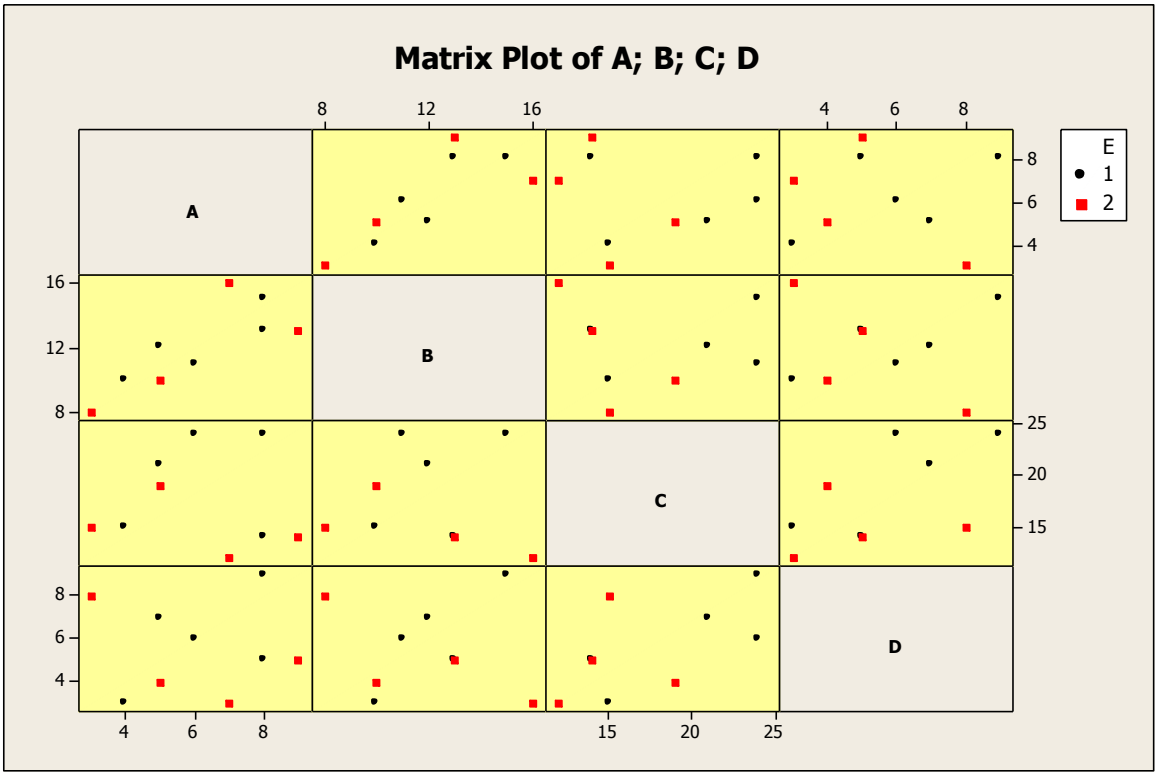
Matrix of plots

This matrix accepts up to 20 variables and creates a plot for every possible combination. A matrix of plots is effective when you have many variables and you would like to see relationships among pairs of variables.



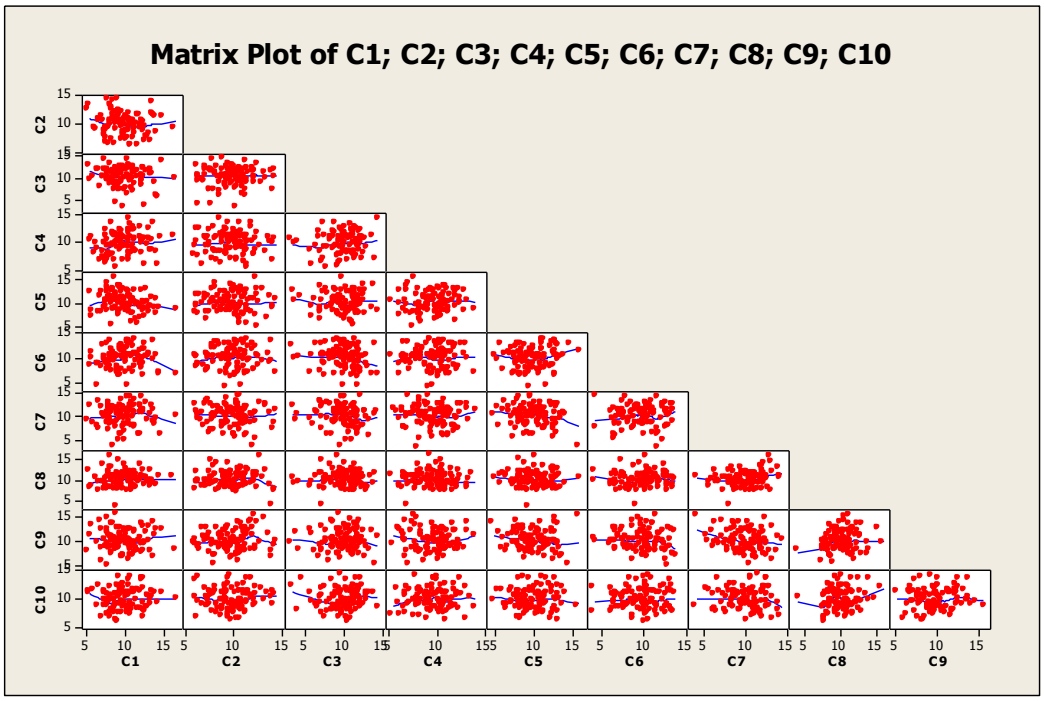
Row	A	B	C	D
1	5	12	21	7
2	7	16	12	3
3	8	13	14	5
4	4	10	15	3
5	3	8	15	8
6	5	10	19	4
7	6	11	24	6
8	8	15	24	9
9	9	13	14	5

Matrix Plot with Groups



Data Display

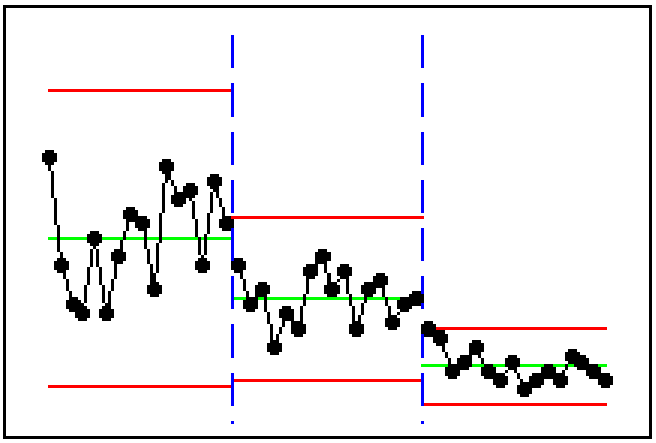
Row	A	B	C	D	E(Category)
1	5	12	21	7	1
2	7	16	12	3	2
3	8	13	14	5	1
4	4	10	15	3	1
5	3	8	15	8	2
6	5	10	19	4	2
7	6	11	24	6	1
8	8	15	24	9	1
9	9	13	14	5	2


Matrix Plot of C1; C2; C3; C4; C5; C6; C7; C8; C9; C10
C2
C3
C4
C5
C6
C7
C8
C9
C10
C1
C2
C3
C4
C5
C6
C7
C8
C9

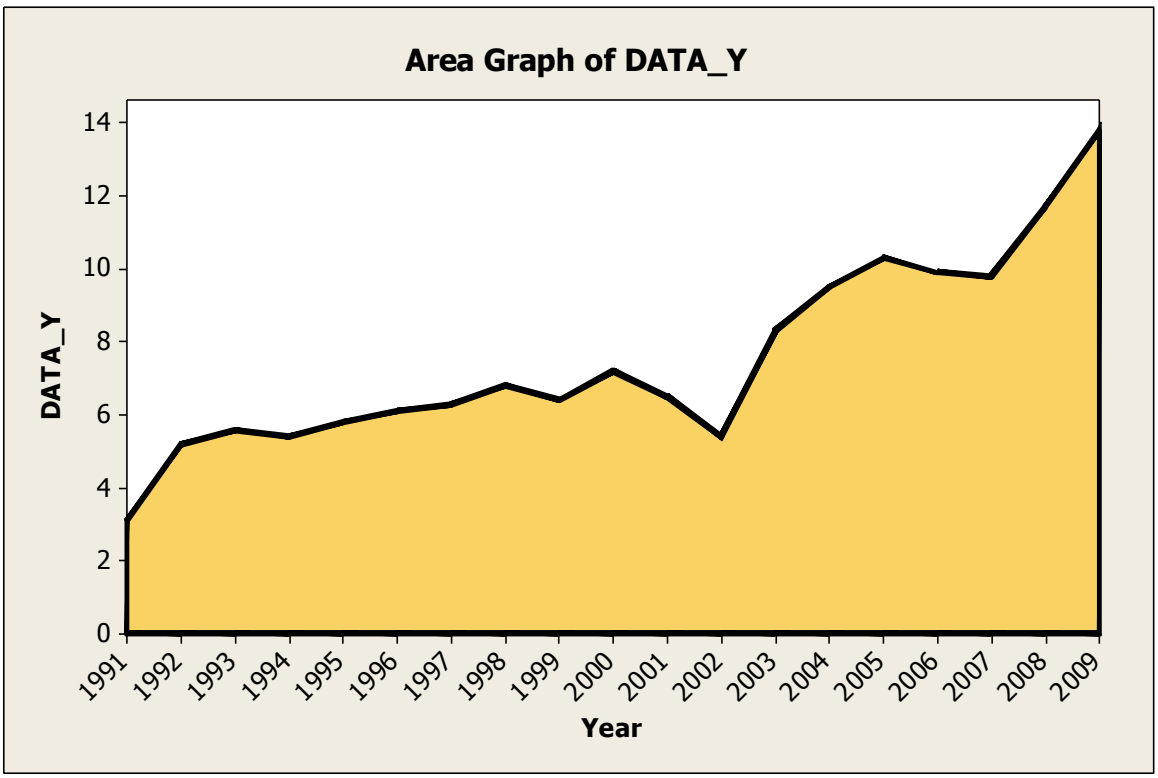
Line Charts

When a quantitative variable is recorded over time at equally spaced intervals (such as daily, weekly, monthly, quarterly, or yearly), the data set forms a time series. Time series data are most effectively presented on a line chart with time as the horizontal axis. The idea is to try to discern a pattern or trend that will likely continue into the future, and then to use that pattern to make accurate predictions for the immediate future.

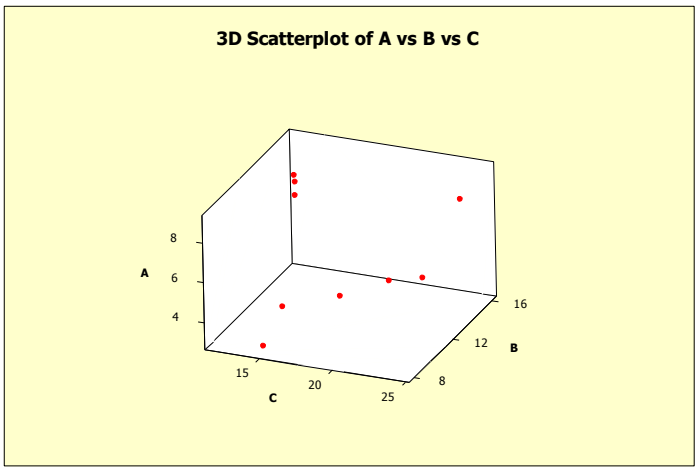
Row	Years	DATA_Y
1	1990	3.1
2	1991	5.2
3	1992	5.6
4	1993	5.4
5	1994	5.8
6	1995	6.1
7	1996	6.3
8	1997	6.8
17	2006	9.8
18	2007	11.7

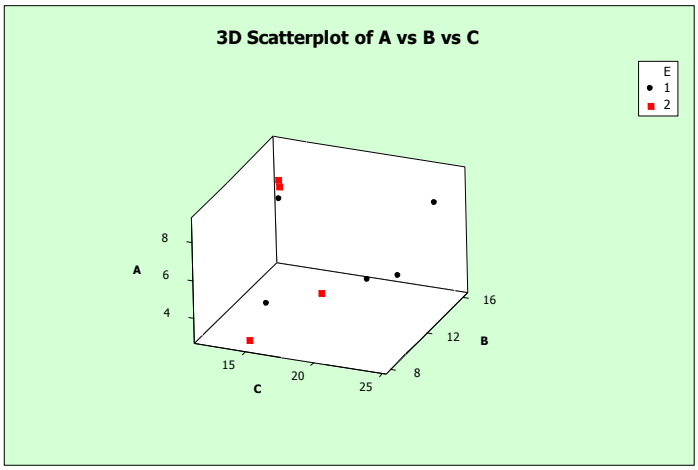


Area Graph


Area Graph of DATA_Y
DATA_Y
0
2
4
6
8
10
12
14
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
Year

3D Scatter Plot

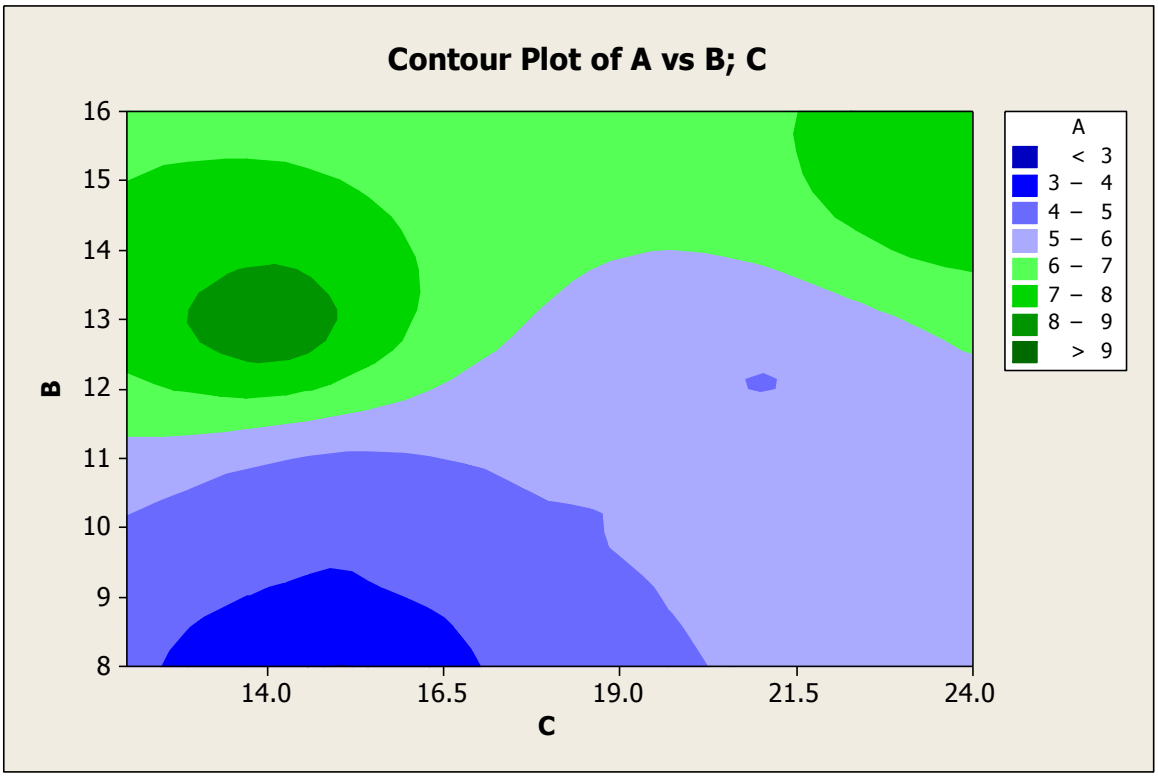




WITH GROUPS

Row	A	B	C	E
1	5	12	21	1
2	7	16	12	2
3	8	13	14	1
4	4	10	15	1
5	3	8	15	2
6	5	10	19	2
7	6	11	24	1
8	8	15	24	1
9	9	13	14	2

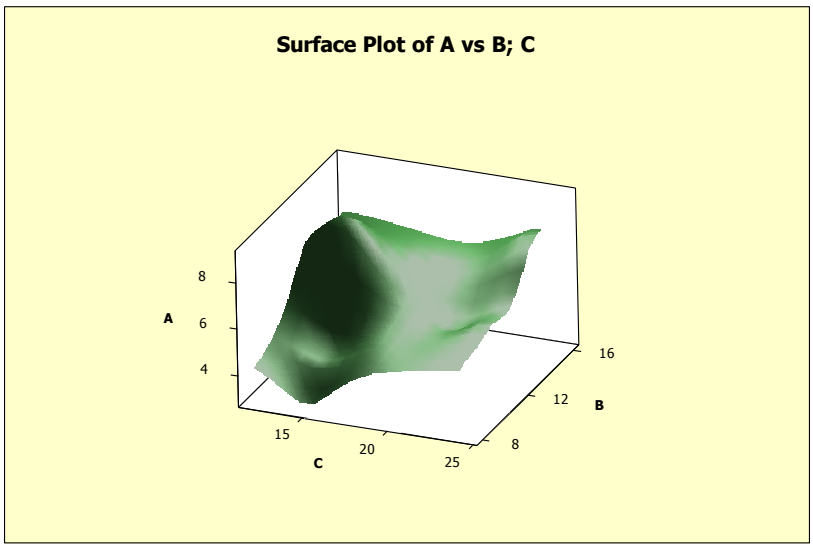
Contour Plot



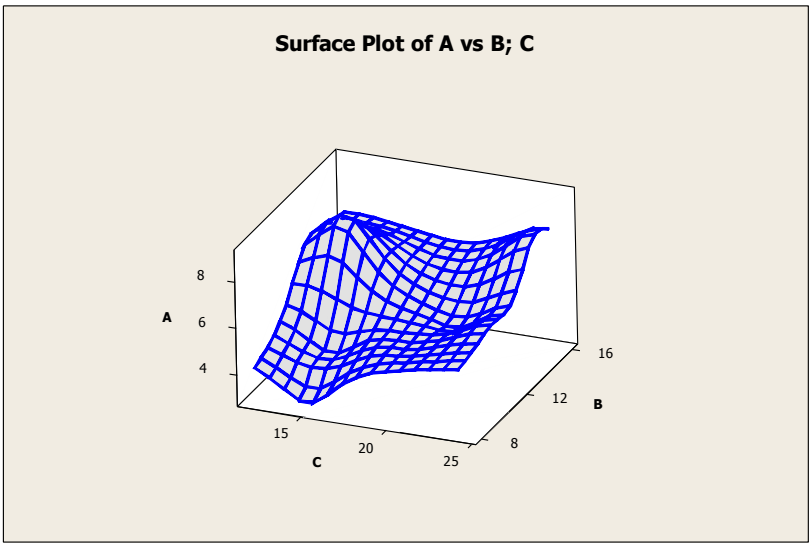
Data Display

Row	A	B	C
1	5	12	21
2	7	16	12
3	8	13	14
4	4	10	15
5	3	8	15
6	5	10	19
7	6	11	24
8	8	15	24
9	9	13	14

Surface Plot



3D Surface PLOT



3D Wireframe PLOT

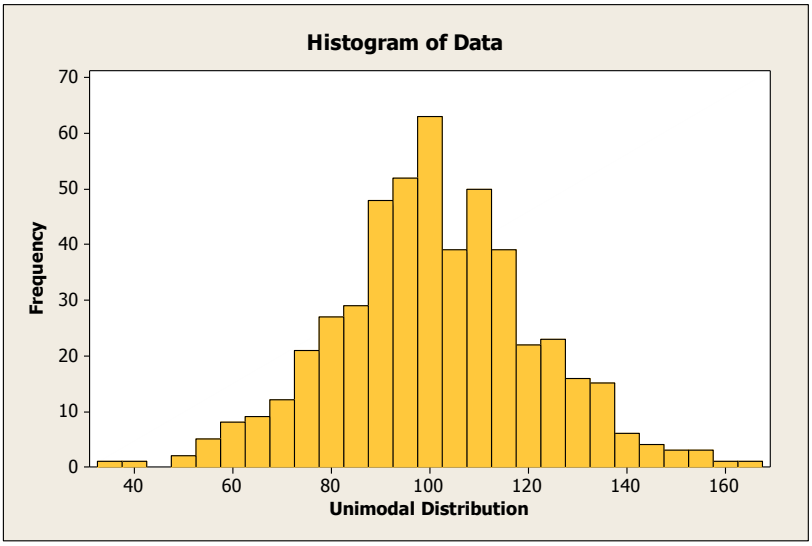
Interpreting Graphs with a Critical Eye

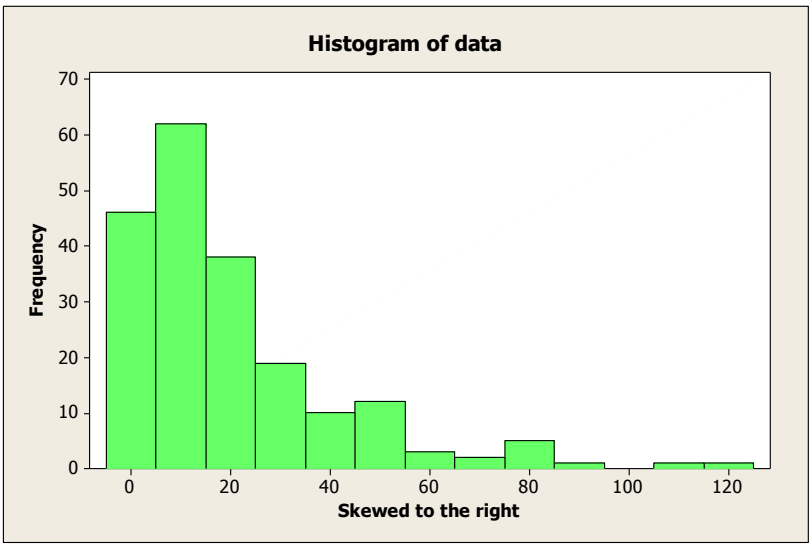
- **First, check the horizontal and vertical scales, so that you are clear about what is being measured.**
- **Examine the location of the data distribution. Where on the horizontal axis is the center of the distribution? If we are comparing two distributions, are they both centered in the same place?**
- **Examine the shape of the distribution. Does the distribution have one peak a point that is higher than any other? If so, this is the most frequently occurring measurement or category. Is there more than one peak? Are there an approximately equal number of measurements to the left and right of the peak?**
- **Look for any unusual measurements or outliers. That is are any measurements much bigger or smaller than all of the others? These outliers may not be representative of the other values in the set.**

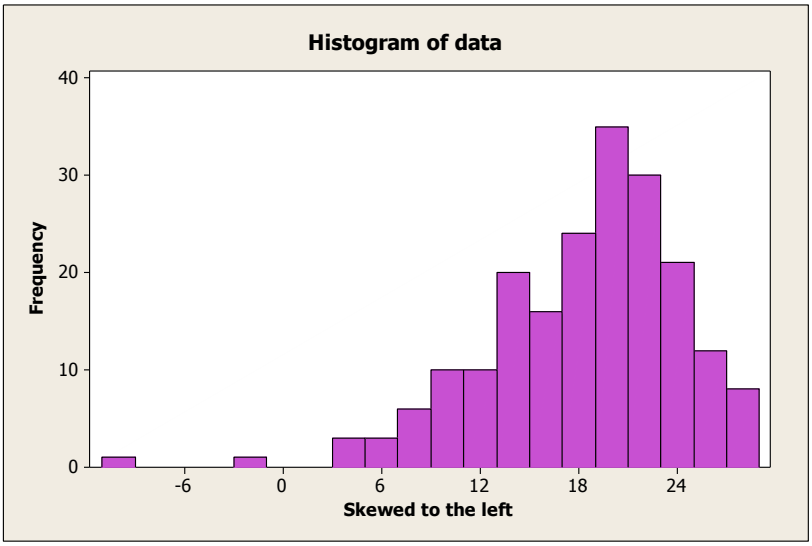
Definition

- ***A distribution is symmetric if the left and right sides of the distribution, when divided at the middle value form mirror images.***
- ***A distribution is skewed to the right if greater proportions of the measurements lie to the right of the peak value. Distributions that are skewed right contain a few unusually large measurements.***
- ***A distribution is skewed to the left if a greater proportion of the measurements lie to the left of the peak value. Distributions that are skewed left contain a few unusually small measurements.***
- ***A distribution is unimodal if it has one peak; a bimodal distribution has two peaks. Bimodal distributions often represent a mixture of two different populations in the data set.***

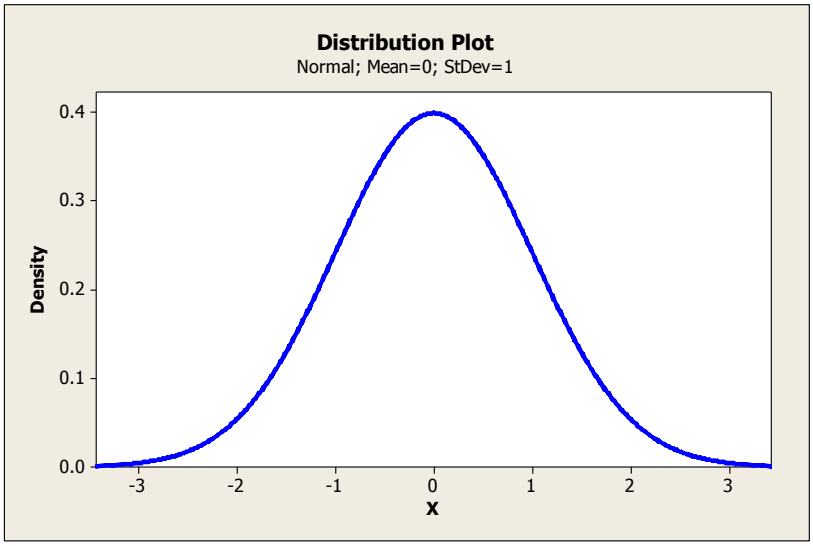
Examine the three histograms.

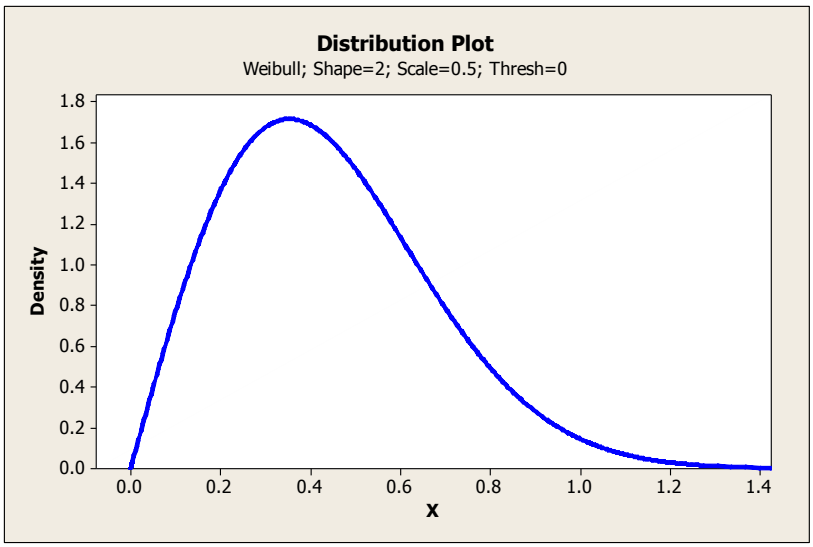


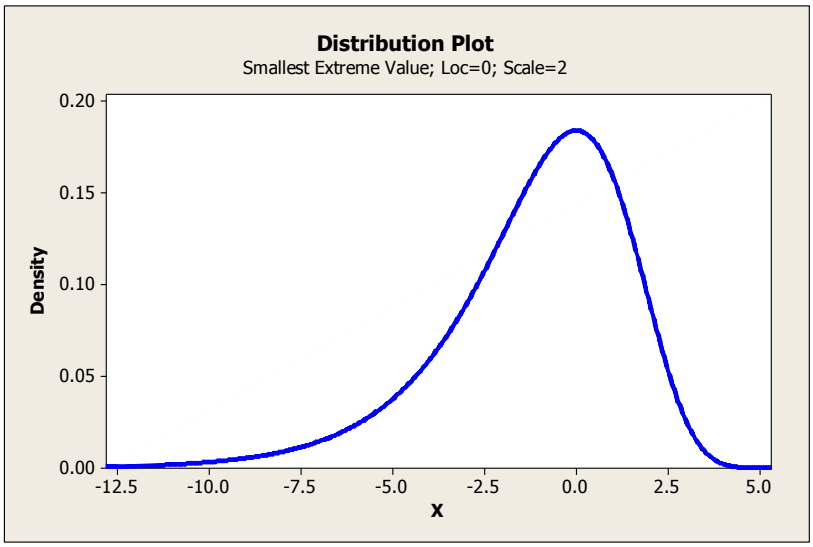




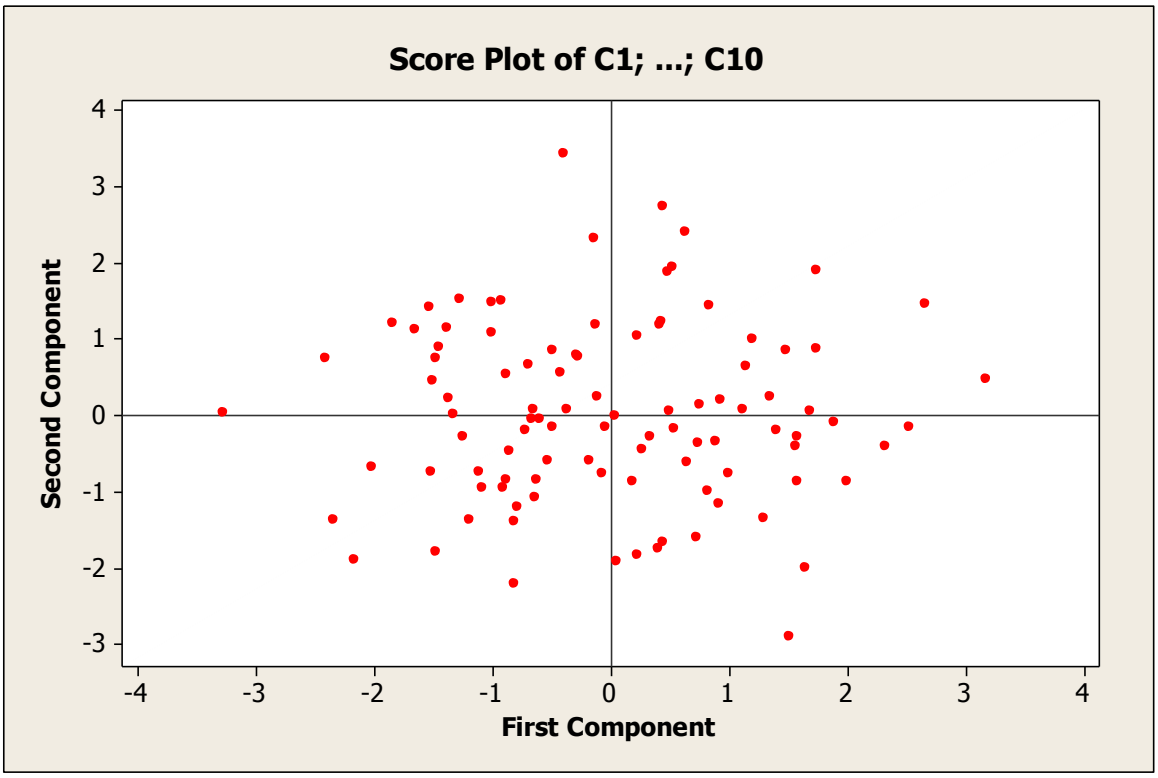
Distribution Plot

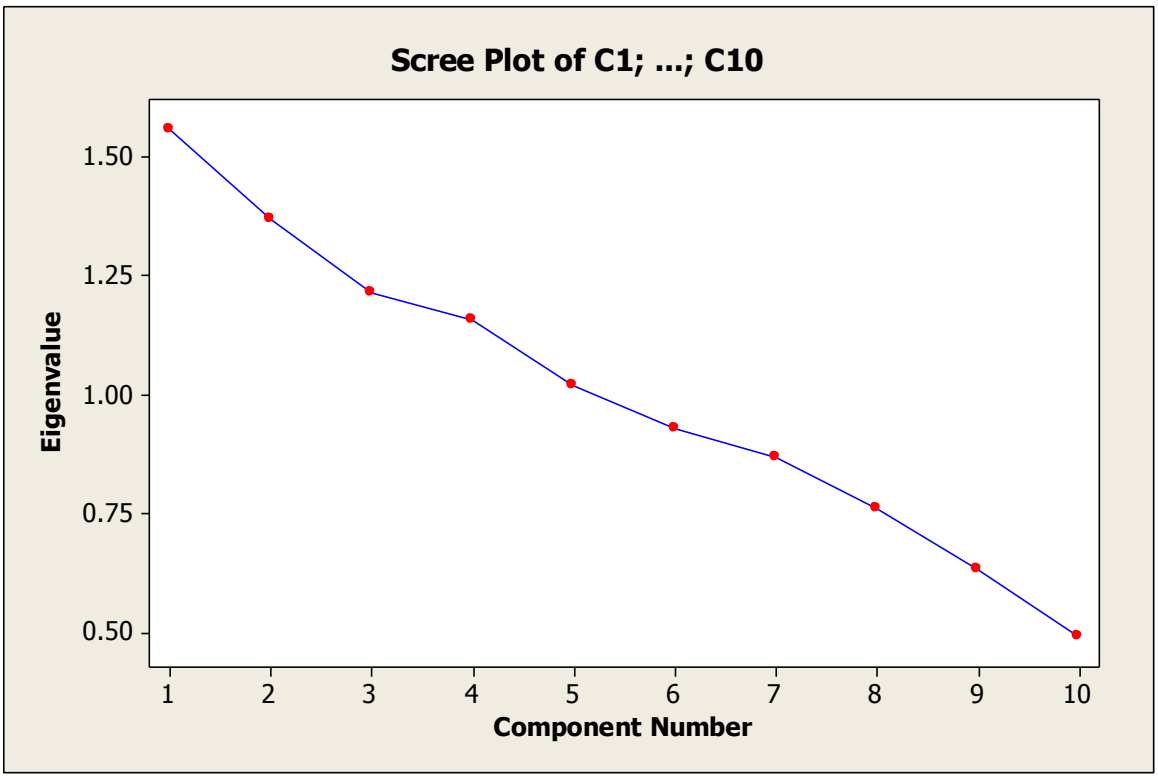


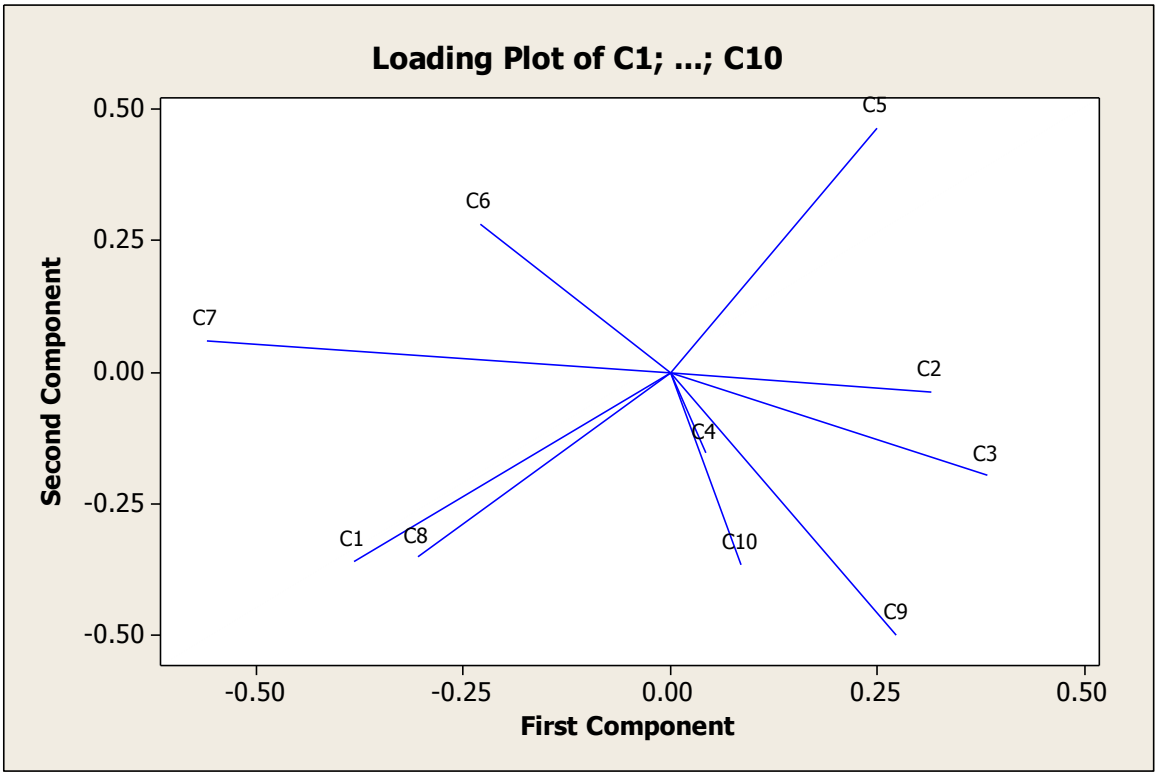


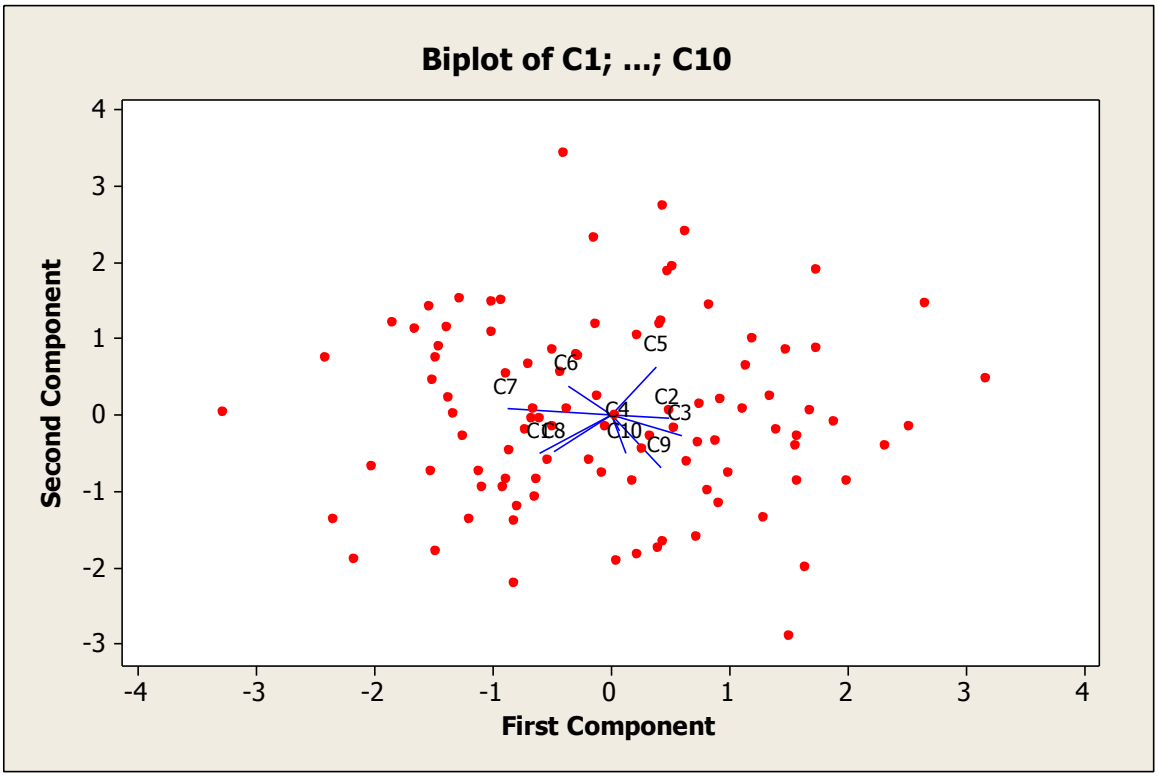


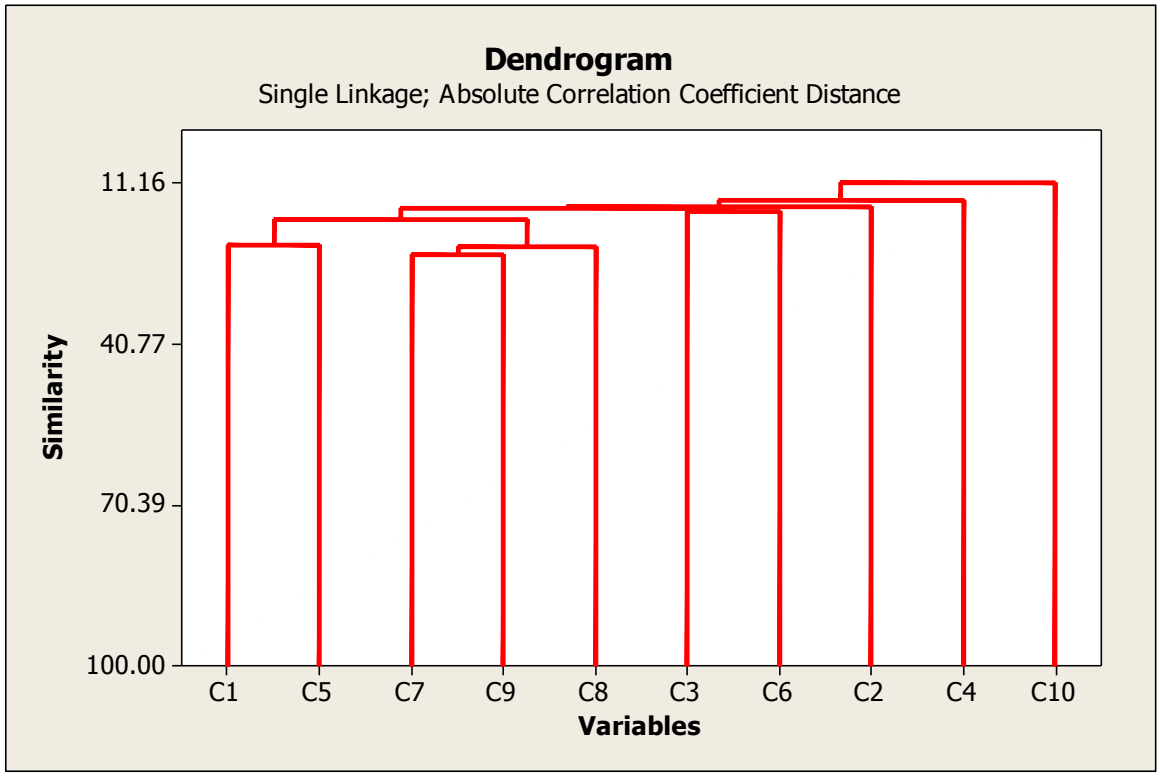
Some Special Plots

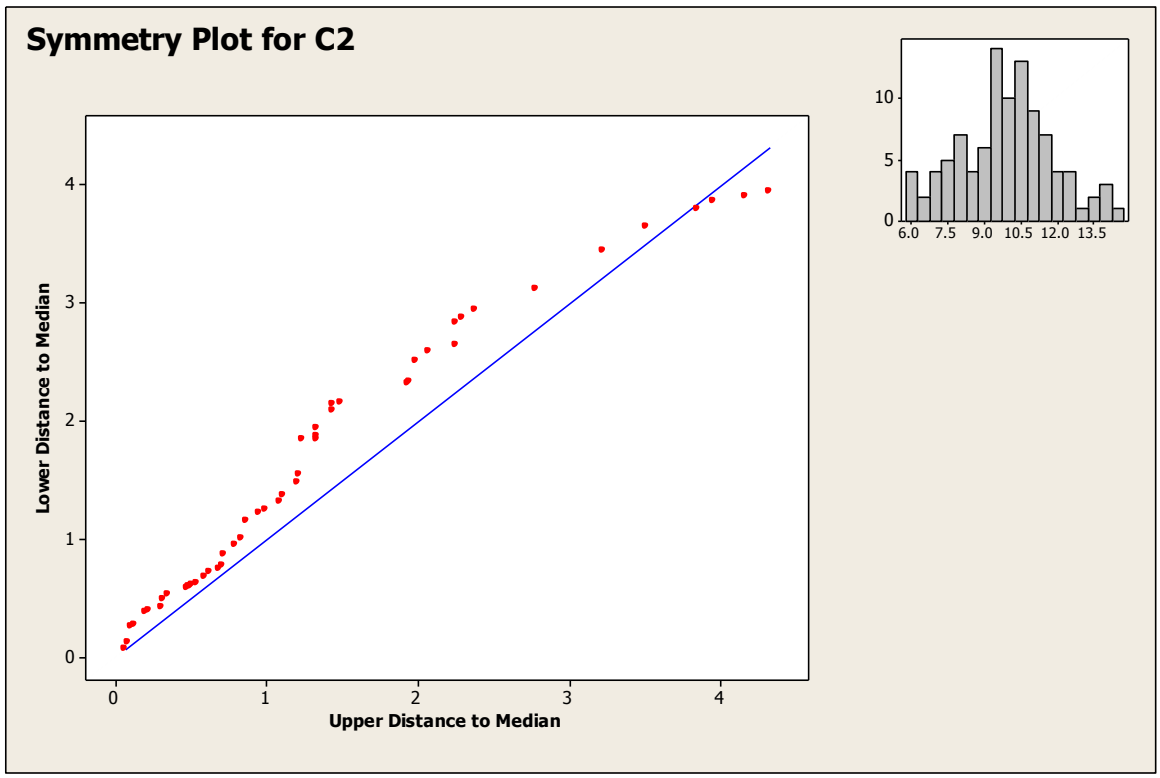




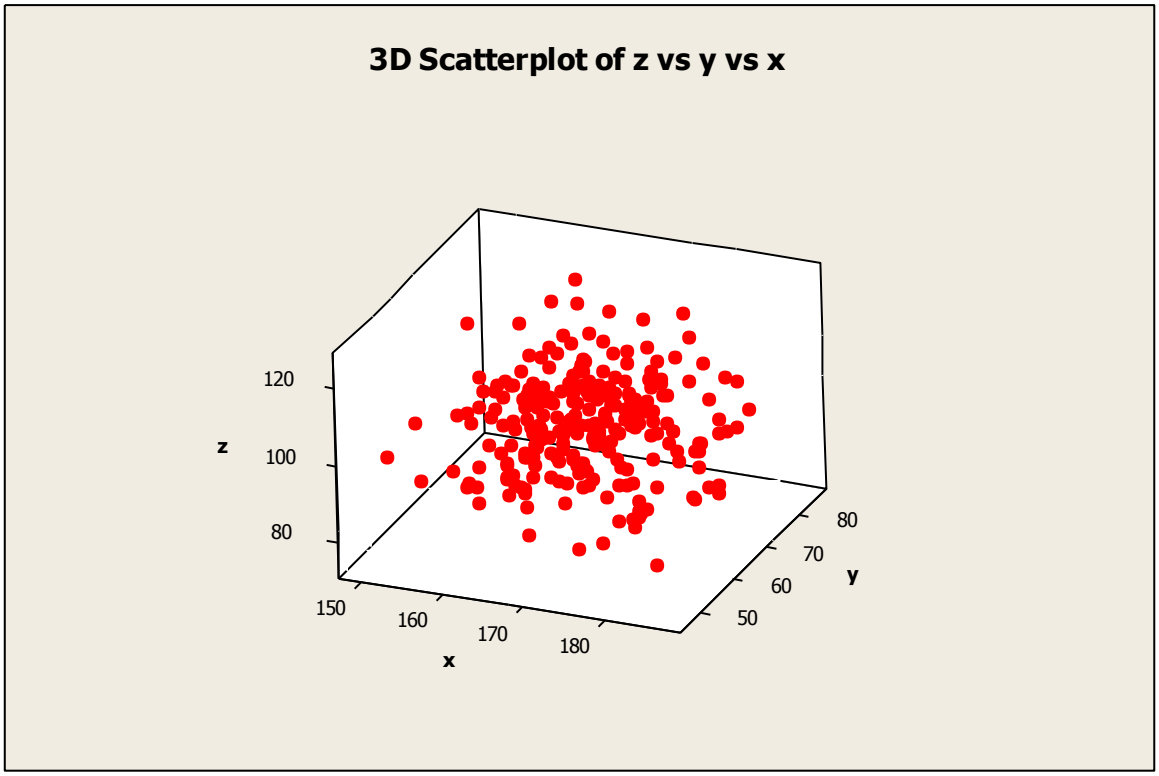

Loading Plot of C1; ...; C10
Second Component
First Component
0.50
0.25
0.00
-0.25
-0.50
-0.50
-0.25
0.00
0.25
0.50
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10


Biplot of C1; ...; C10
Second Component
First Component
4
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10

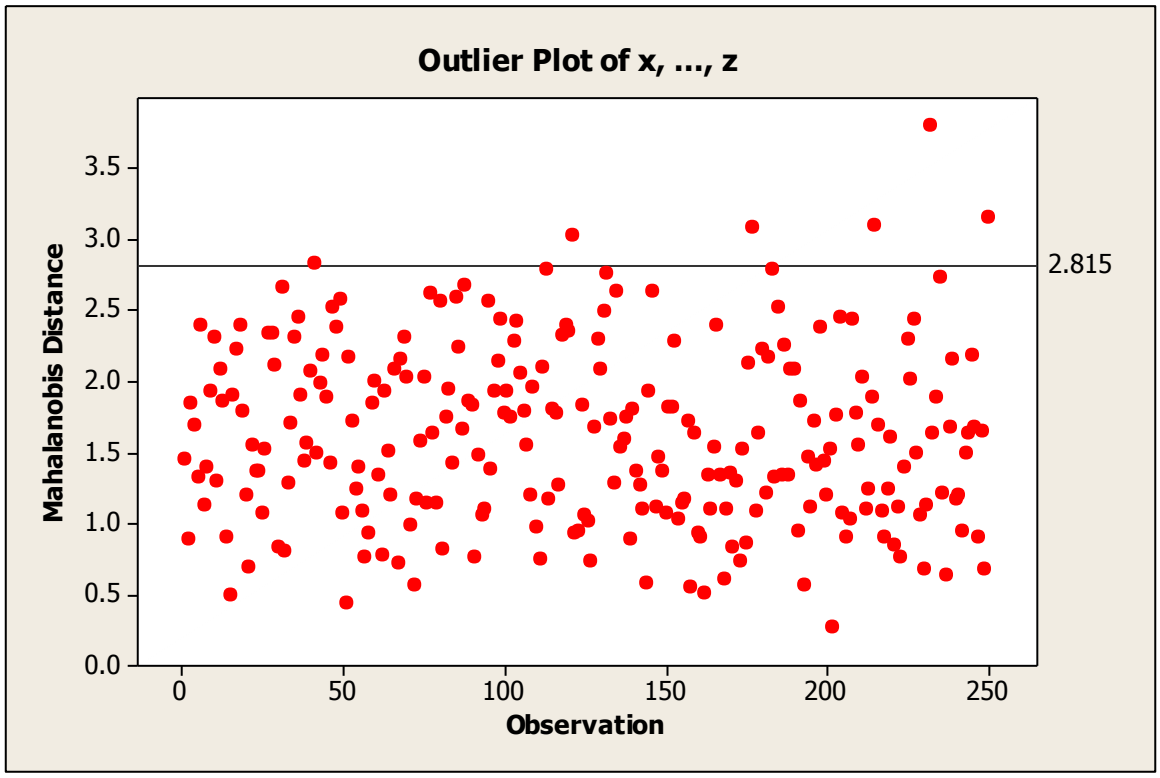

Dendrogram
Single Linkage; Absolute Correlation Coefficient Distance
Similarity
11.16
40.77
70.39
100.00
C1
C5
C7
C9
C8
C3
C6
C2
C4
C10
Variables


Symmetry Plot for C2
Lower Distance to Median
Upper Distance to Median
0
1
2
3
4
10
5
6.0
7.5
9.0
10.5
12.0
13.5

3D Plot


3D Scatterplot of z vs y vs x
z
120
100
80
150
160
170
180
x
80
70
60
50
y

Outlier Plot



Mahalanobis distance

The Mahalanobis distance measures the distance from each point in multivariate space to the overall mean or centroid, utilizing the covariance structure of the data.