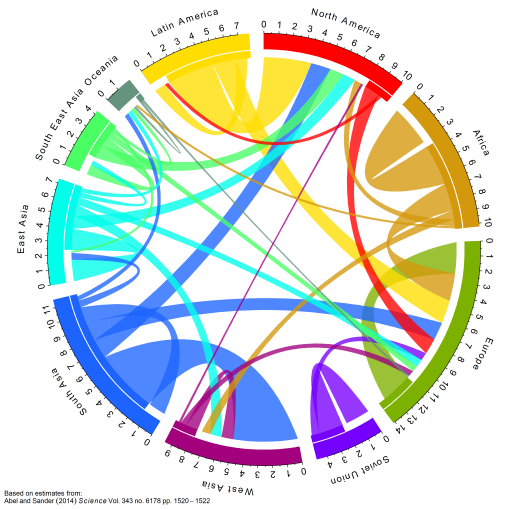
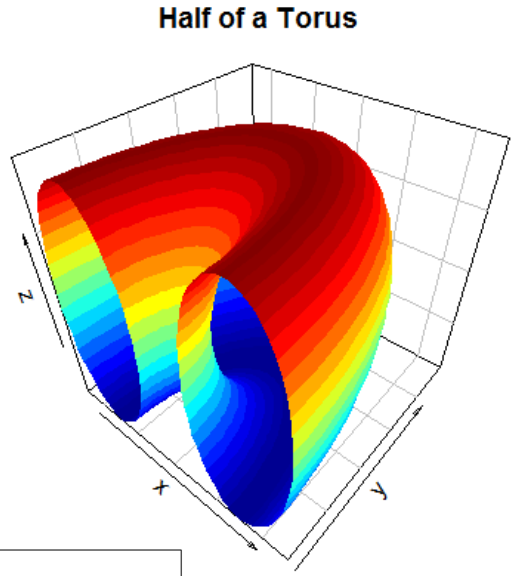
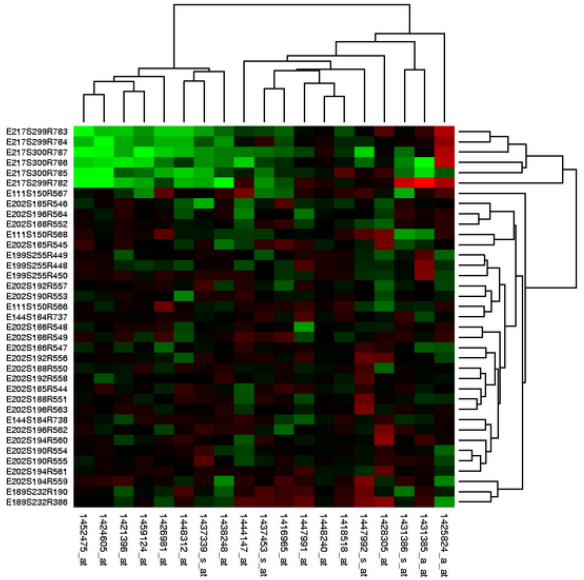
VISUALIZATION

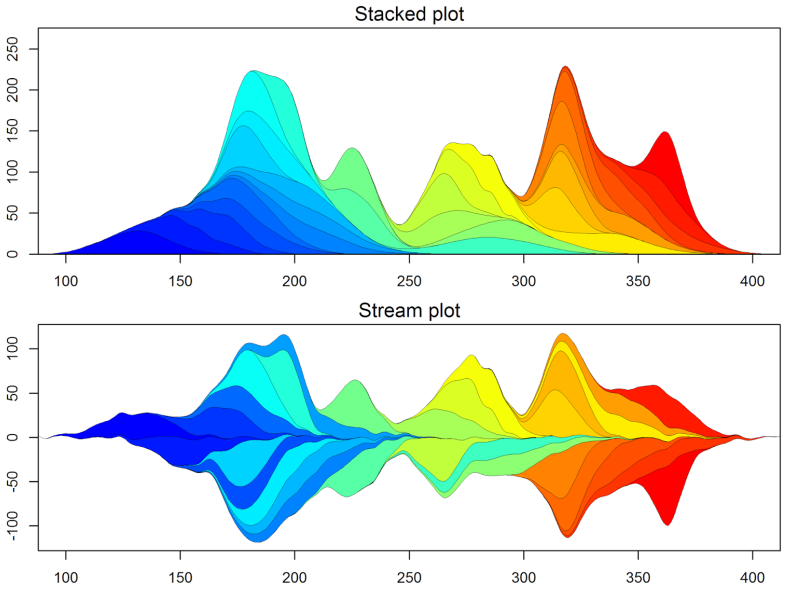
Introducción a la Ciencia de Datos

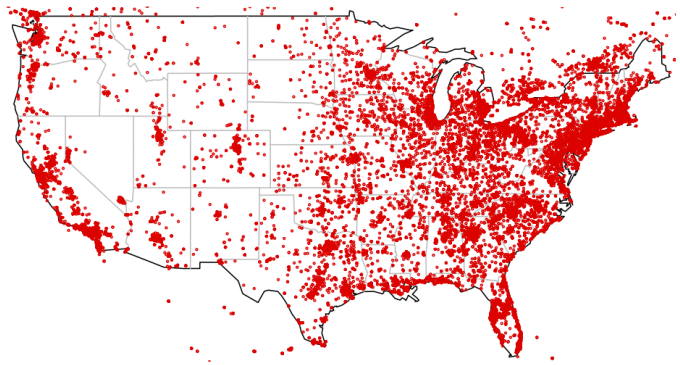
Gallery of plots









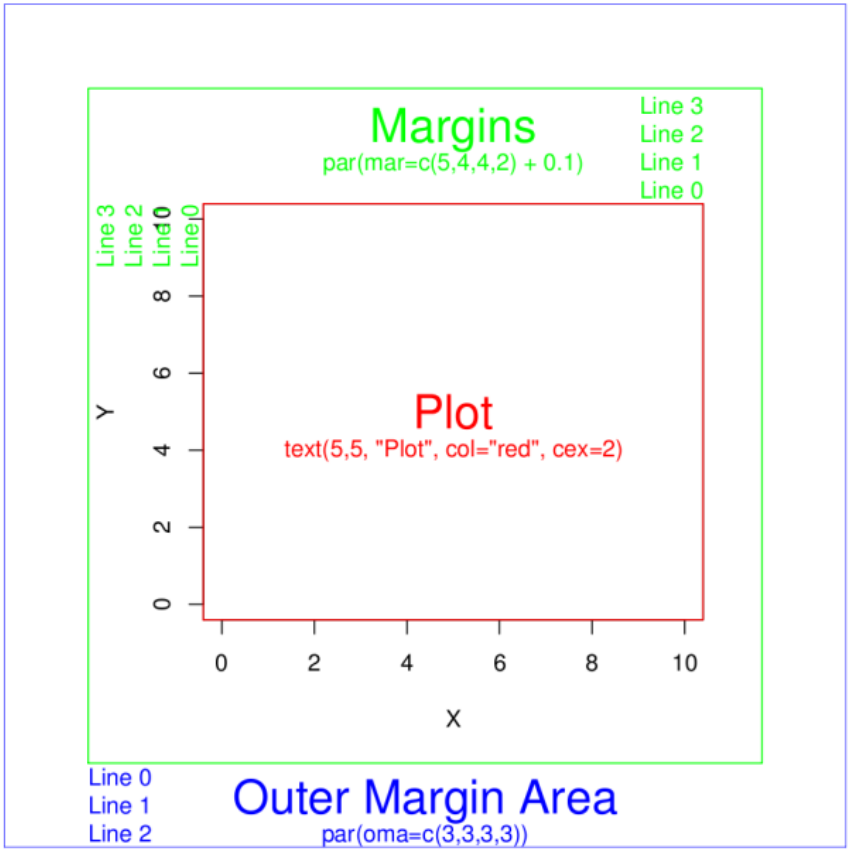


R plots

- We will see a small set of them:
  - Plots
  - Histograms
  - Boxplots
  - Barplots
  - Pies

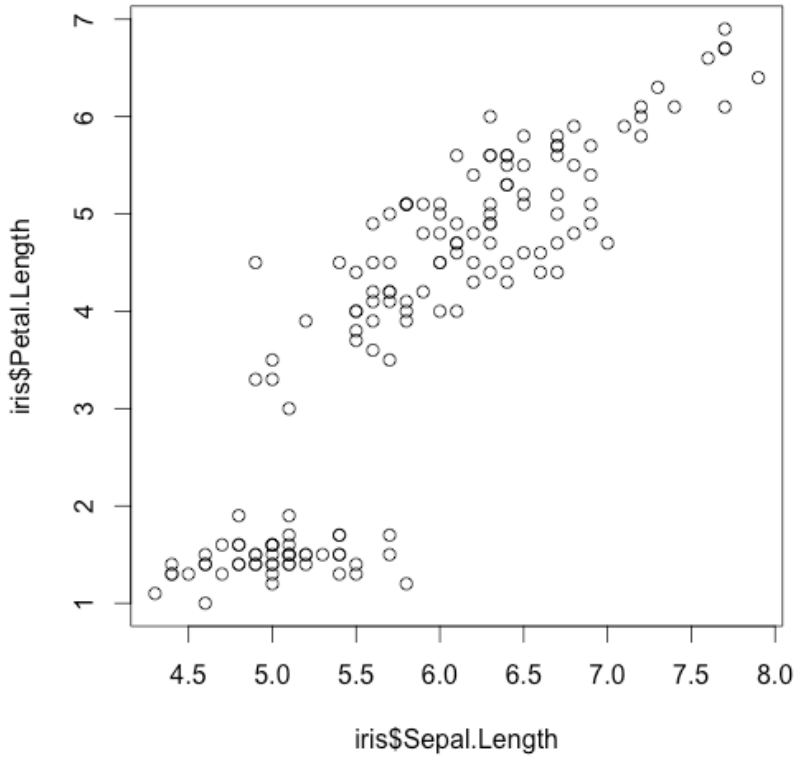
Plot

- To change parameters for margins and plot areas use the `par` function.
- Link to R manual



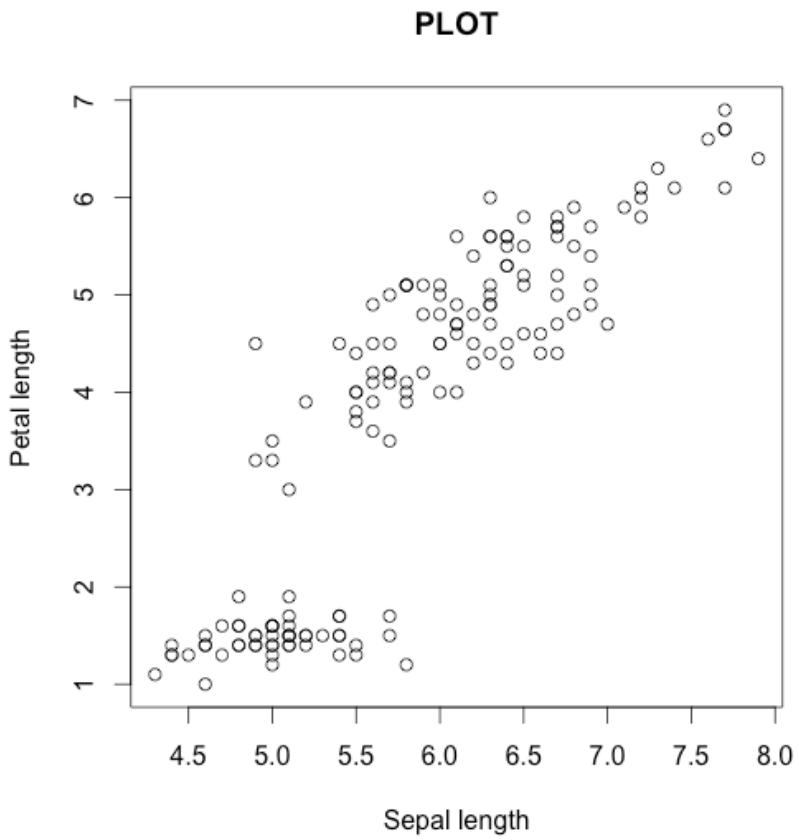
Plot

- `?iris`
- `iris`
- `class(iris)`
- `names(iris)`
- `?plot`
- `example(plot)`
- `plot(iris$Sepal.Length, iris$Petal.Length)`



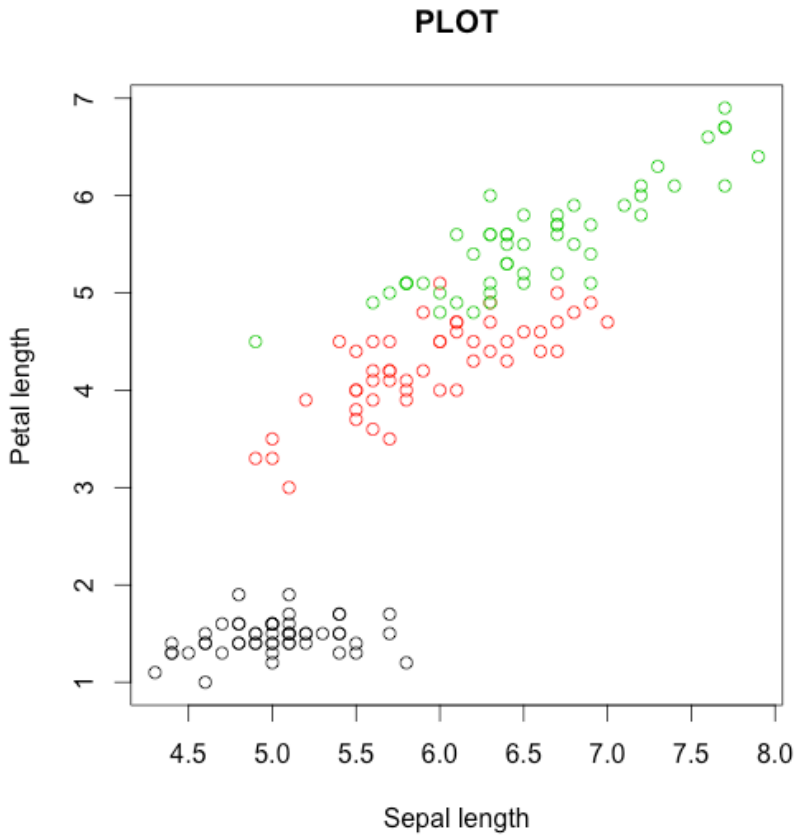
Plot

- `plot(iris$Sepal.Length, iris$Petal.Length, main="PLOT", xlab="Sepal length", ylab="Petal length")`



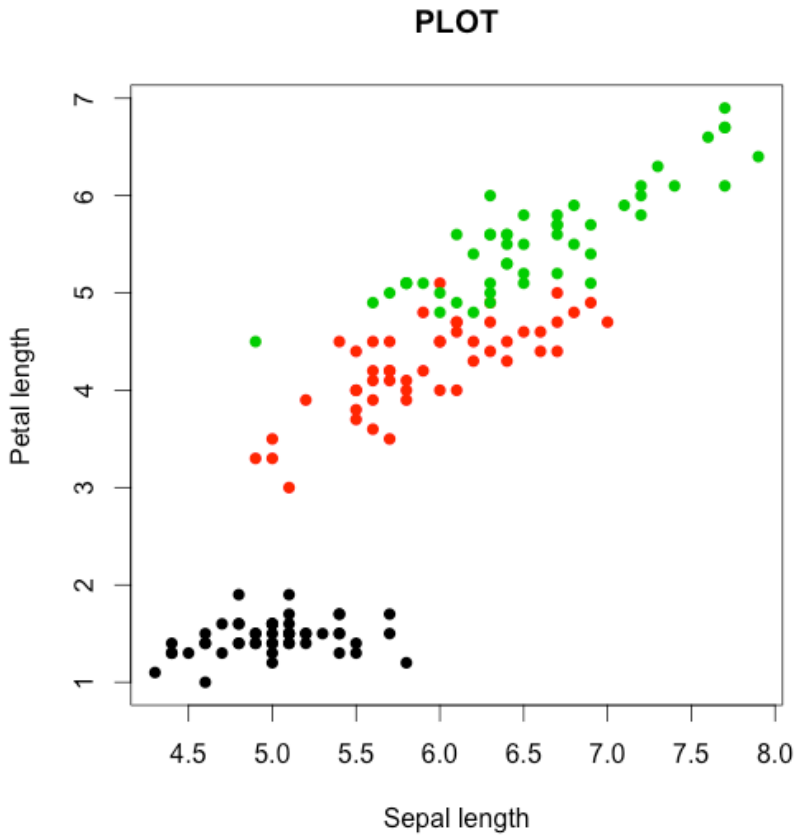
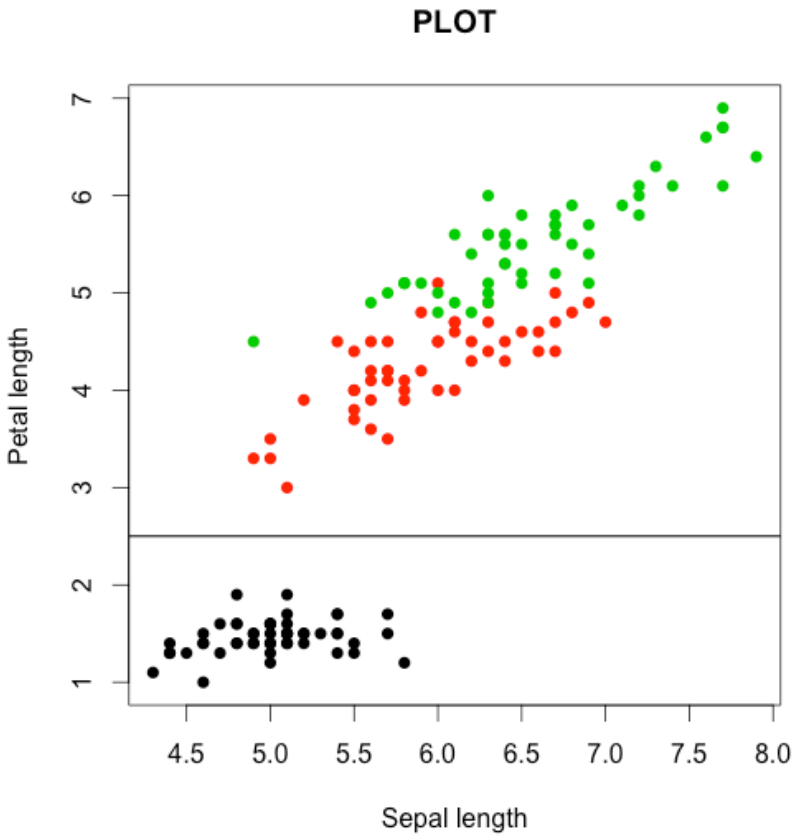
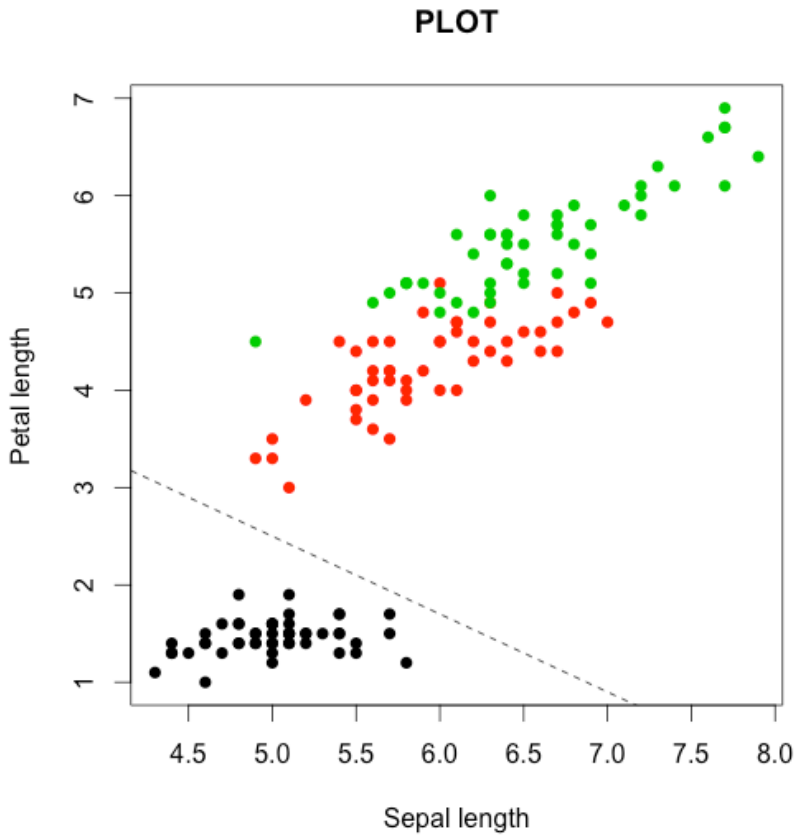
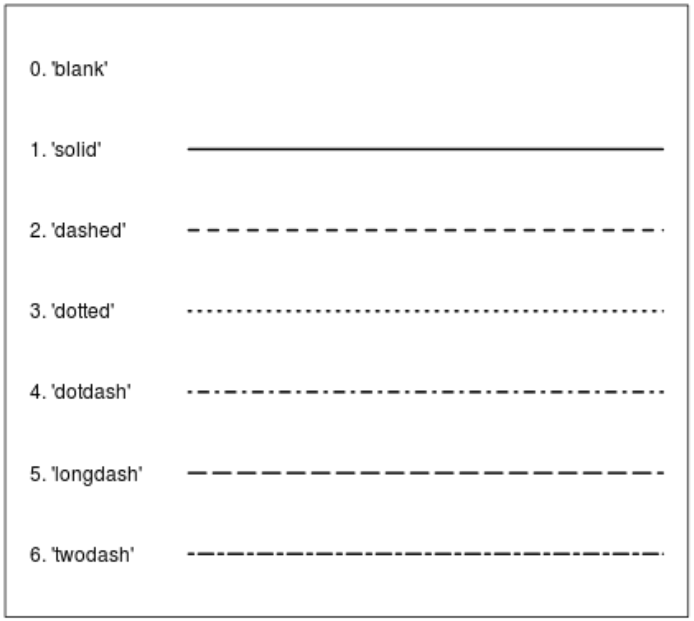
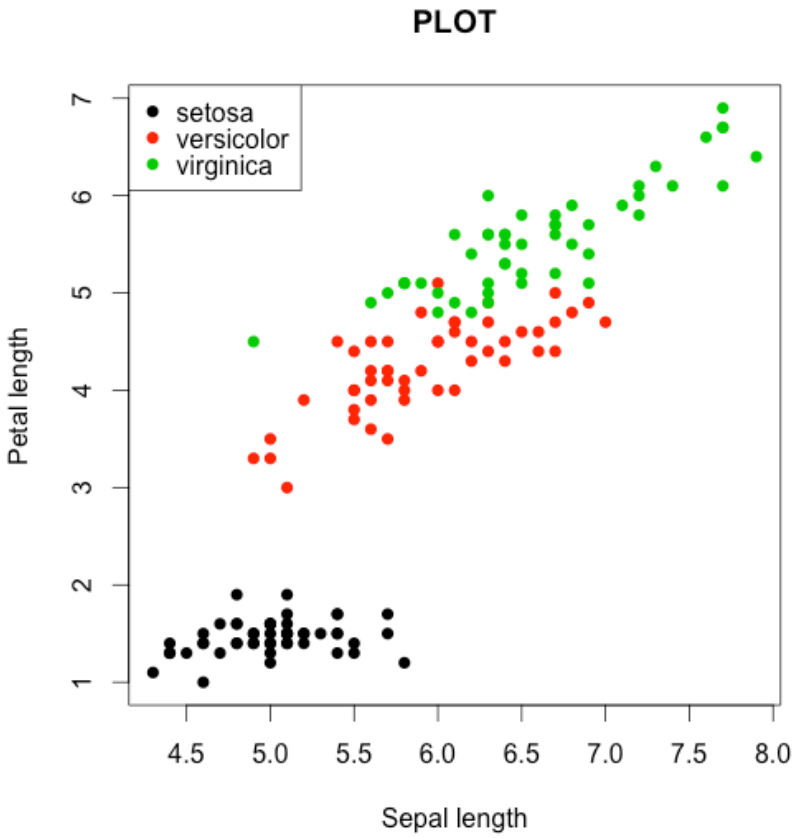
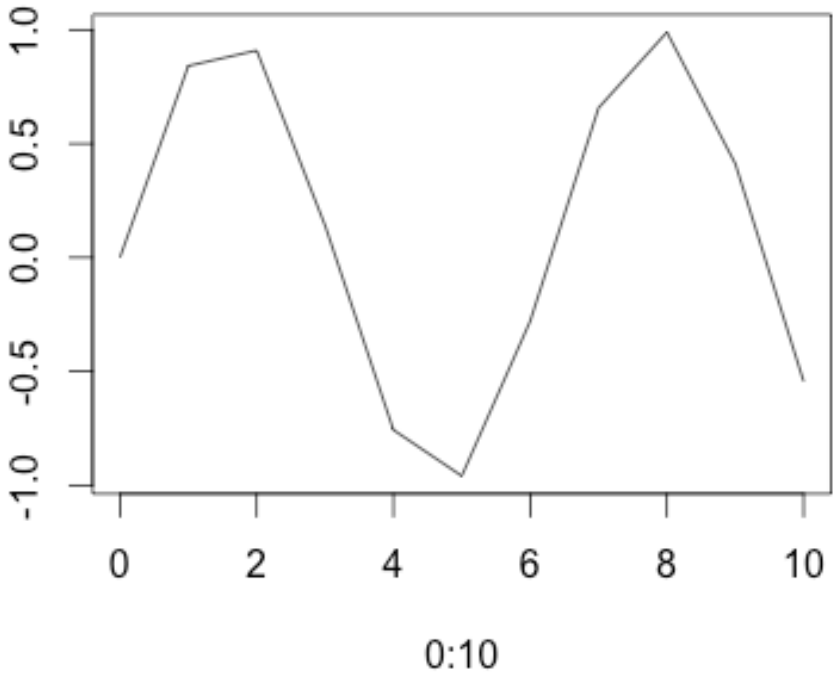
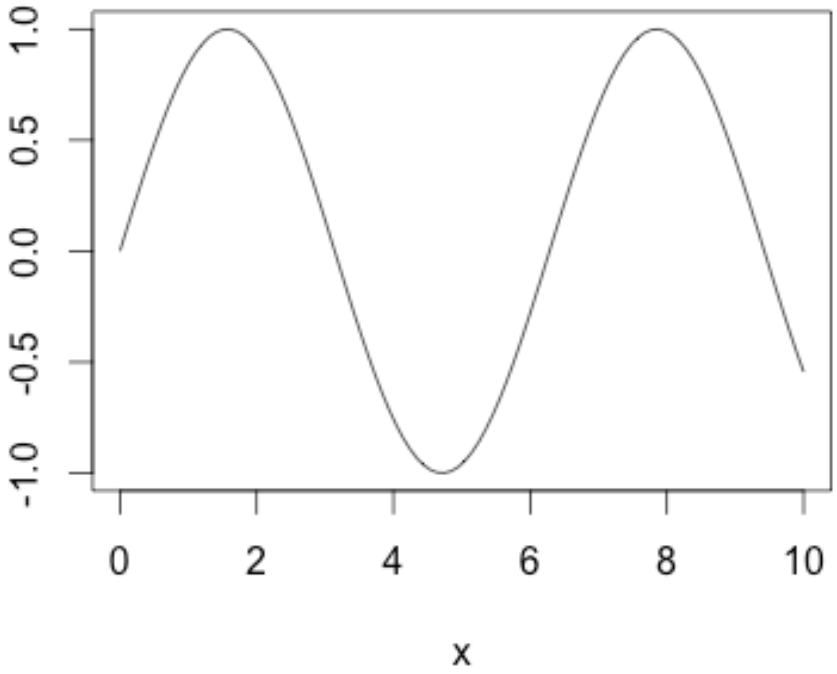
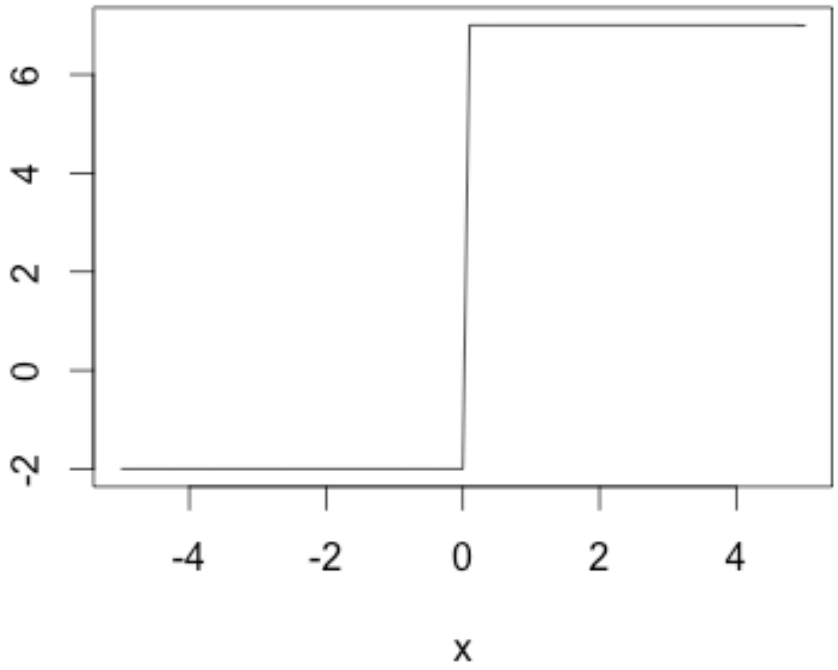
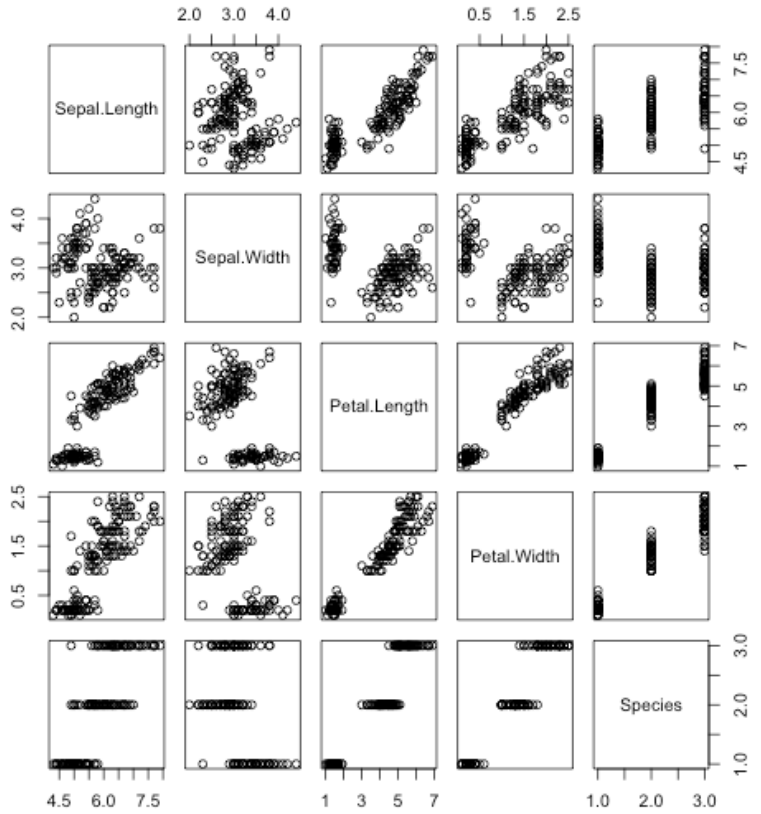
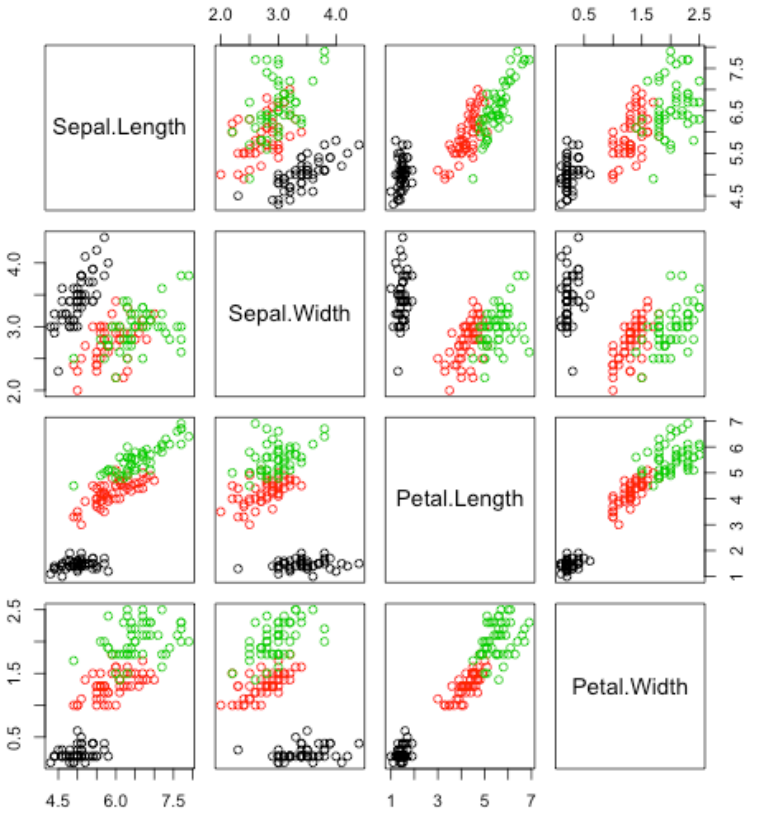
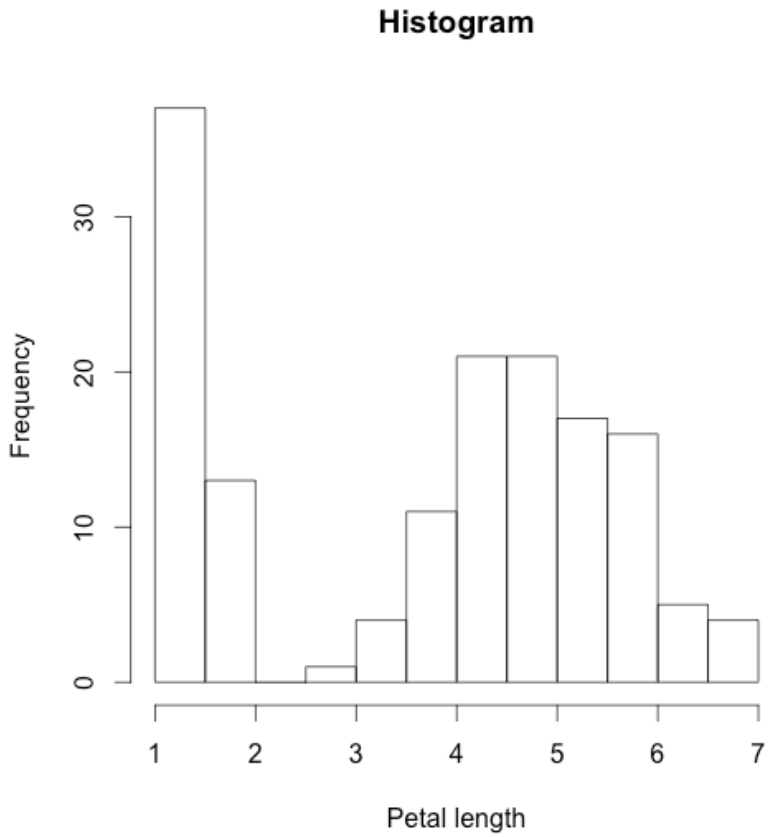
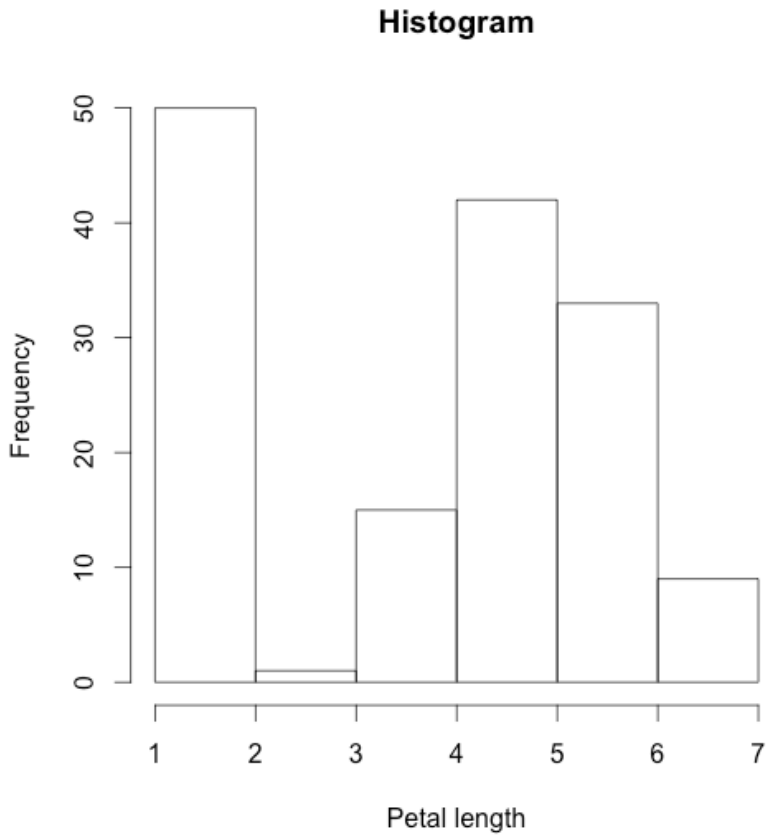
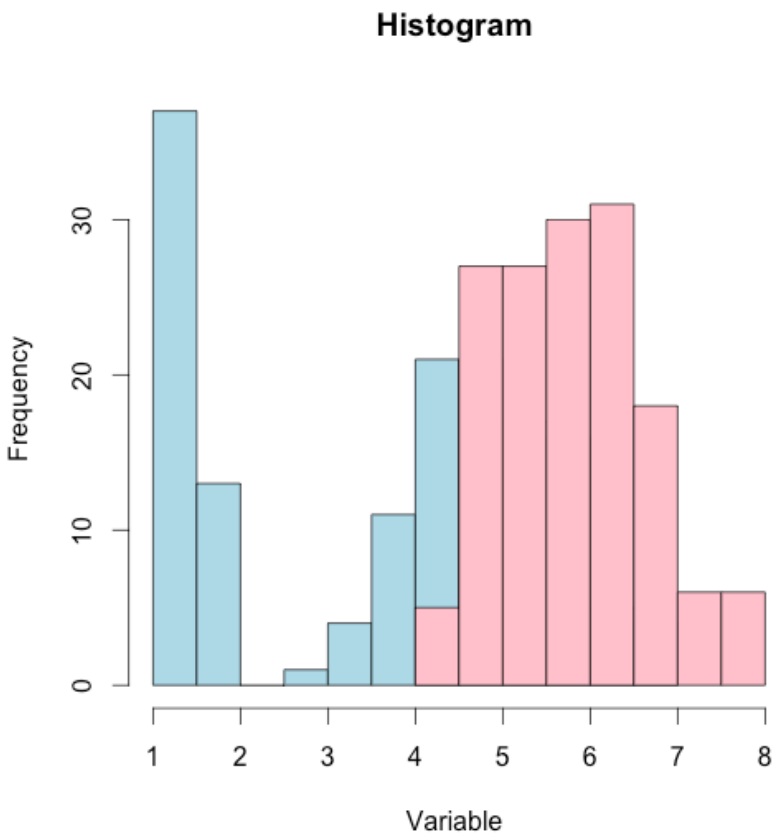
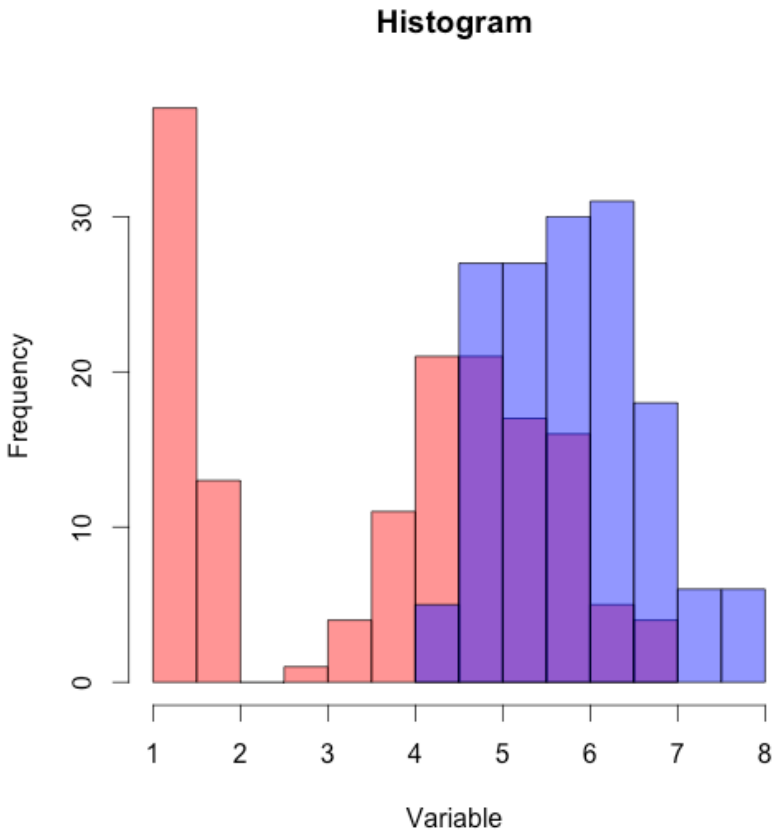
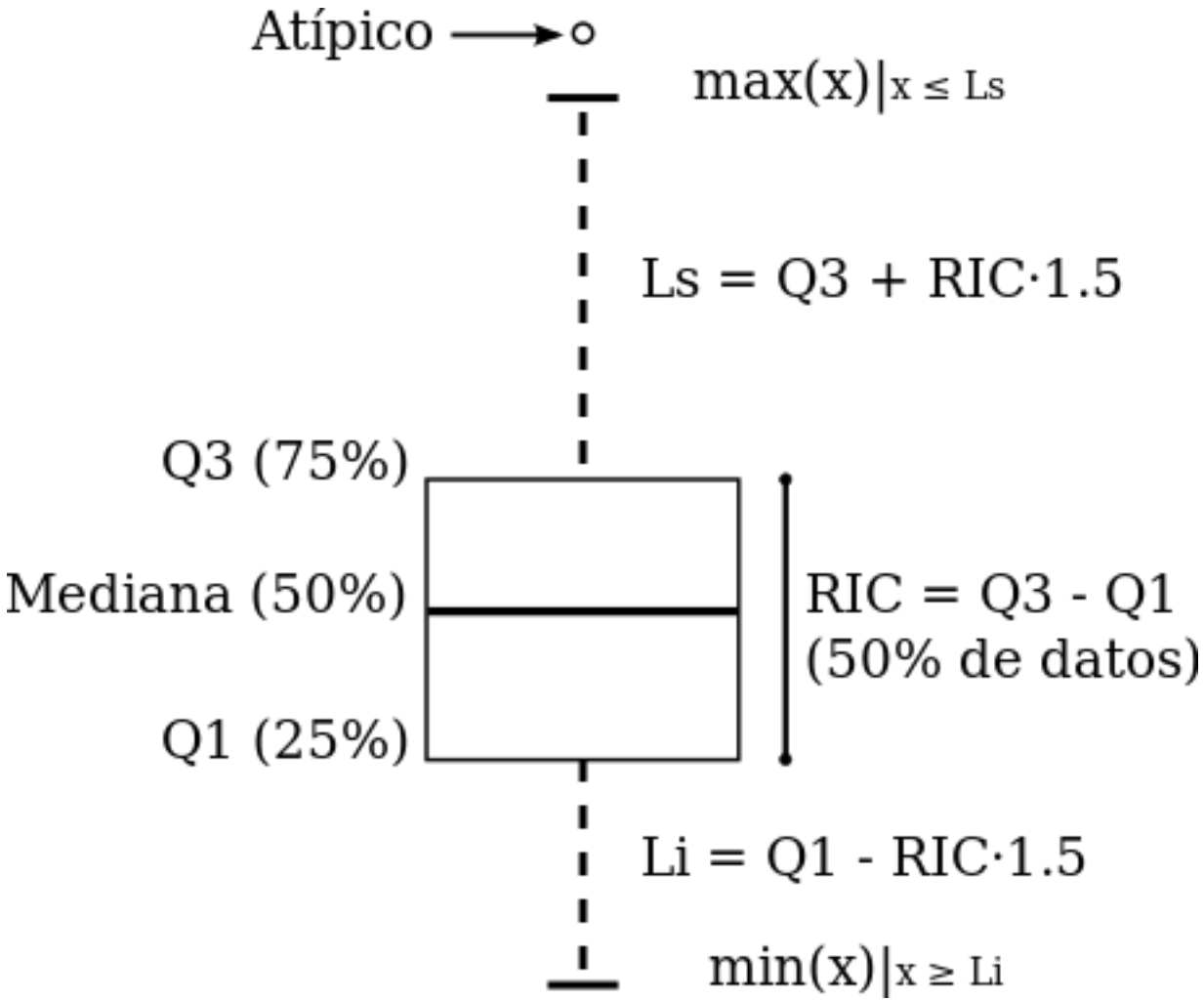
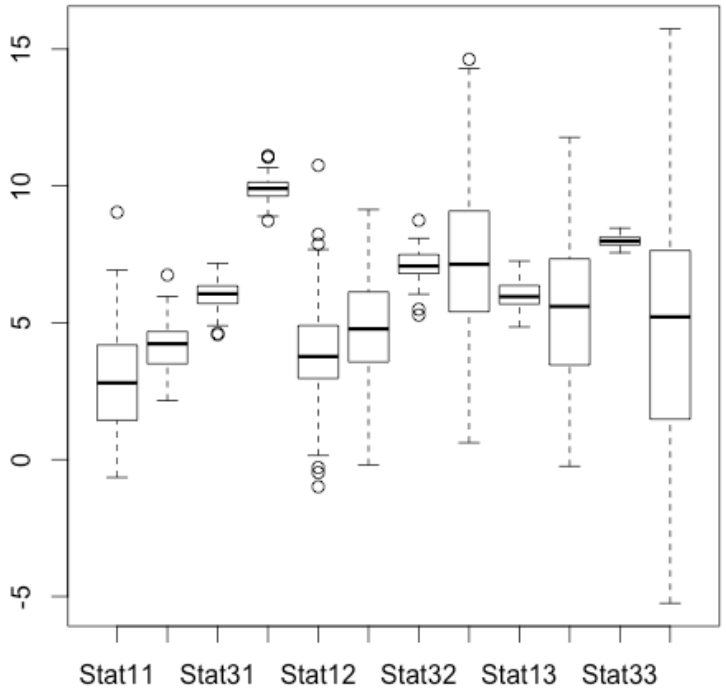
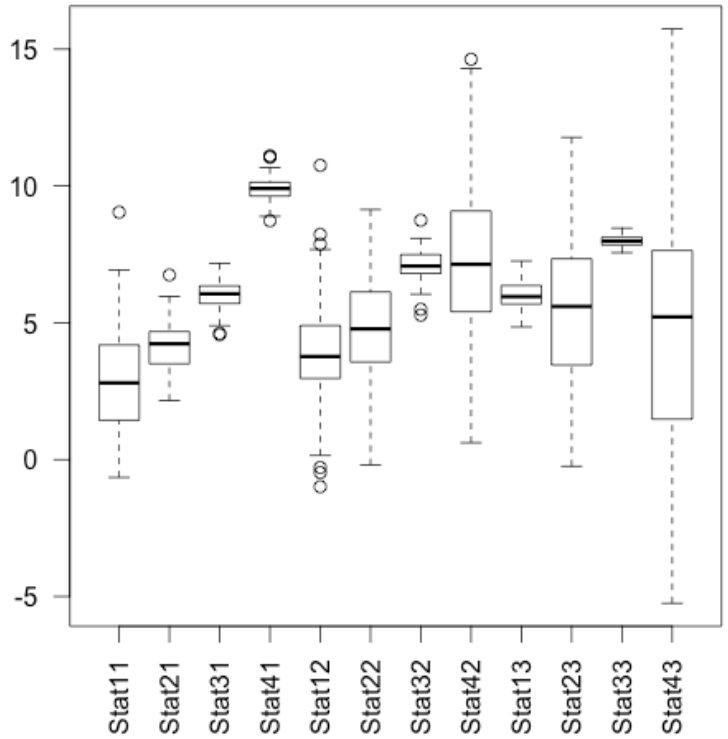
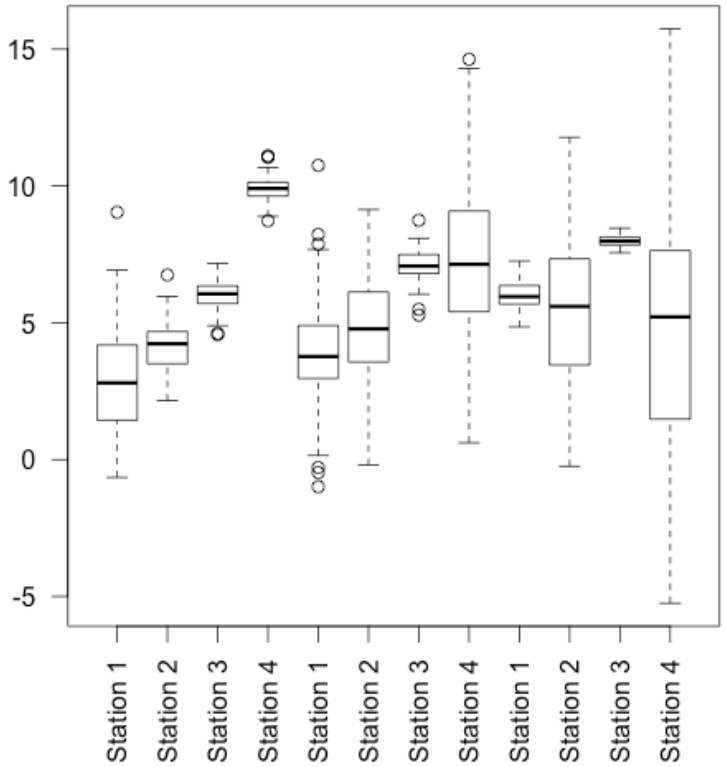
Plot

- `plot(iris$Sepal.Length, iris$Petal.Length, col=iris$Species, main="PLOT", xlab="Sepal length", ylab="Petal length")`



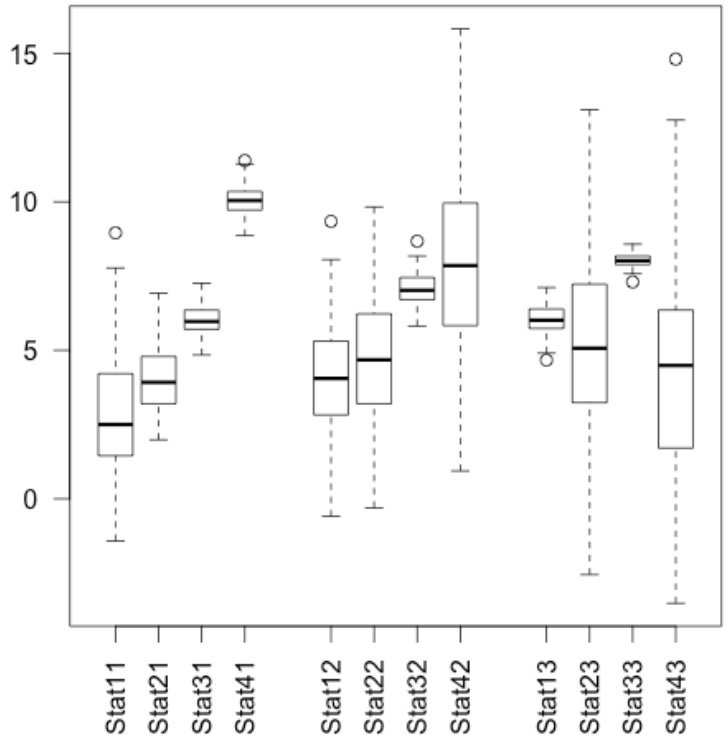
Plot

- ```
  plot(iris$Sepal.Length,
  pch=16,
  iris$Petal.Length,
  col=iris$Species,
  main="PLOT", xlab="Sepal
  length", ylab="Petal
  length")
  ```


  ```

Plot

pch


0:
10:
20:
A: A
1:
11:
21:
a: a
2:
12:
22:
B: B
3:
13:
23:
b: b
4:
14:
24:
S: S
5:
15:
25:
`: `
6:
16:
@: @
.: .
7:
17:
+: +
,: ,
8:
18:
%: %
?: ?
9:
19:
#: #
*: *

Plot

- `plot(iris$Sepal.Length, pch=16, iris$Petal.Length, col=iris$Species, main="PLOT", xlab="Sepal length", ylab="Petal length")`
- `?abline`
- `abline(h=2.5)`



Plot

- `plot(iris$Sepal.Length, pch=16, iris$Petal.Length, col=iris$Species, main="PLOT", xlab="Sepal length", ylab="Petal length")`
- `abline(6.5, -0.8, lty=2)`



Plot

pch


0:
10:
20:
A: A
1:
11:
21:
a: a
2:
12:
22:
B: B
3:
13:
23:
b: b
4:
14:
24:
S: S
5:
15:
25:
`: `
6:
16:
@: @
.: .
7:
17:
+: +
,: ,
8:
18:
%: %
?: ?
9:
19:
#: #
*: *

lty


0. 'blank'
1. 'solid'
2. 'dashed'
3. 'dotted'
4. 'dotdash'
5. 'longdash'
6. 'twodash'

Plot

- `plot(iris$Sepal.Length, pch=16, iris$Petal.Length, col=iris$Species, main="PLOT", xlab="Sepal length", ylab="Petal length")`
- `legend("topleft", legend=levels(iris$Species), col=1:length(levels(iris$Species)), pch=16)`



Curves

- `plot(0:10, sin(0:10), type="l”)`



Curve

- `curve(sin(x), 0, 10, n=100)`



Piecewise functions

- `myfun <- function(x) {ifelse(x > 0,7,-2)}`
- `curve(myfun(x), -5, 5)`



Scatterplot

- `plot(iris)`



Scatterplot

- `plot(iris[,-5], col=iris$Species)`



Histogram

- `?hist`
- `hist(iris$Petal.Length, main="Histogram", xlab="Petal length")`



Histogram

- `hist(iris$Petal.Length, main="Histogram", xlab="Petal length", breaks=5)`



Histogram

- `minimo <- min(iris$Petal.Length,iris$Sepal.Length)`
- `maximo <- max(iris$Petal.Length,iris$Sepal.Length)`
- `hist(iris$Petal.Length, main="Histogram", xlab="Variable", col="lightblue", xlim=c(minimo,maximo))`
- `hist(iris$Sepal.Length, main="Histogram", xlab="Variable", col="pink", add=TRUE)`



Histogram

- `hist(iris$Petal.Length, main="Histogram", xlab="Variable", col=rgb(1,0,0,0.5), xlim=c(minimo,maximo))`
- `hist(iris$Sepal.Length, main="Histogram", xlab="Variable", col=rgb(0,0,1,0.5), add=TRUE)`



Boxplot


Atípico
max(x)|x ≤ Ls
Ls = Q3 + RIC·1.5
Q3 (75%)
Mediana (50%)
Q1 (25%)
RIC = Q3 - Q1
(50% de datos)
Li = Q1 - RIC·1.5
min(x)|x ≥ Li

Boxplot

```
data<-
data.frame(Stat11=rnorm(100,mean=3,
sd=2),
Stat21=rnorm(100,mean=4,sd=1),
Stat31=rnorm(100,mean=6,sd=0.5),
Stat41=rnorm(100,mean=10,sd=0.5),
Stat12=rnorm(100,mean=4,sd=2),
Stat22=rnorm(100,mean=4.5,sd=2),
Stat32=rnorm(100,mean=7,sd=0.5),
Stat42=rnorm(100,mean=8,sd=3),
Stat13=rnorm(100,mean=6,sd=0.5),
Stat23=rnorm(100,mean=5,sd=3),
Stat33=rnorm(100,mean=8,sd=0.2),
Stat43=rnorm(100,mean=4,sd=4))
```

- `summary(data)`
- `boxplot(data)`



Boxplot

- `boxplot(data, las = 2)`



Boxplot

- ```
  boxplot(data, las = 2,
  names = c("Station
  1","Station 2","Station
  3","Station 4","Station
  1","Station 2","Station
  3","Station 4","Station
  1","Station 2","Station
  3","Station 4"))
  ```


  ```

Boxplot

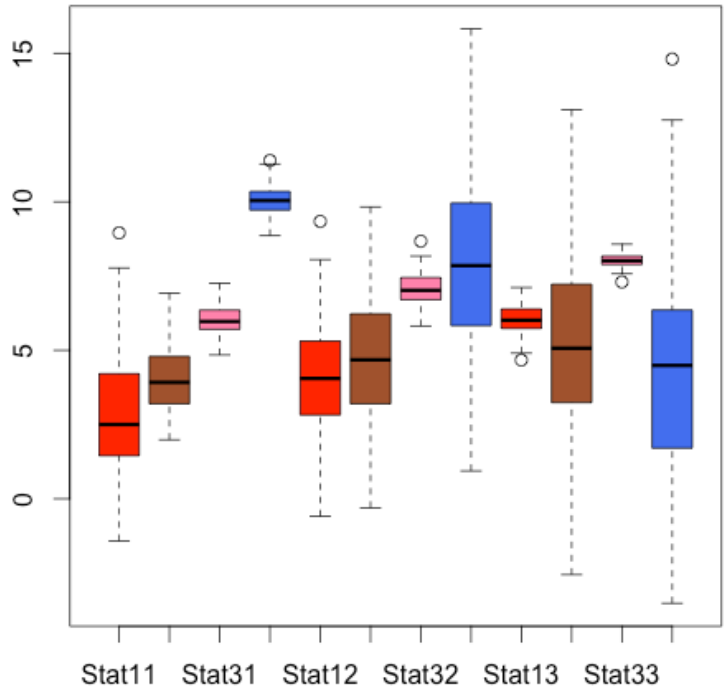
- `boxplot(data, las = 2, at = c(1,2,3,4,6,7,8,9,11,12, 13,14))`



Boxplot

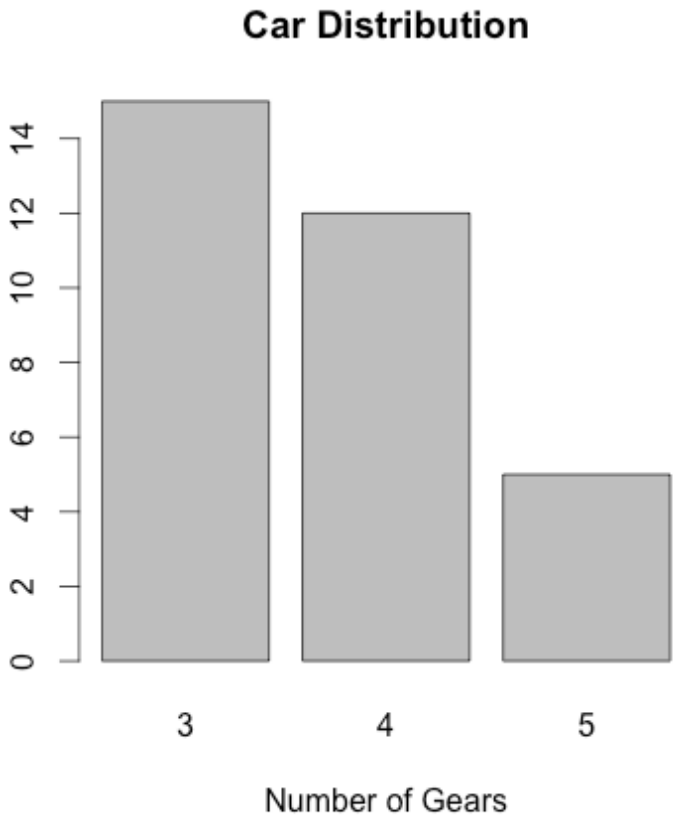
- `boxplot(data, col = c("red","sienna","palevioletred1","royalblue2","red","sienna","palevioletred1","royalblue2","red","sienna","palevioletred1","royalblue2"))`

- `colors()`



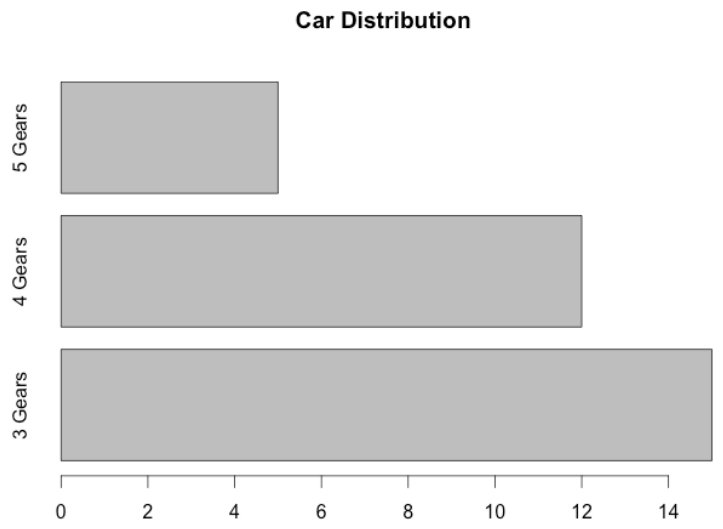
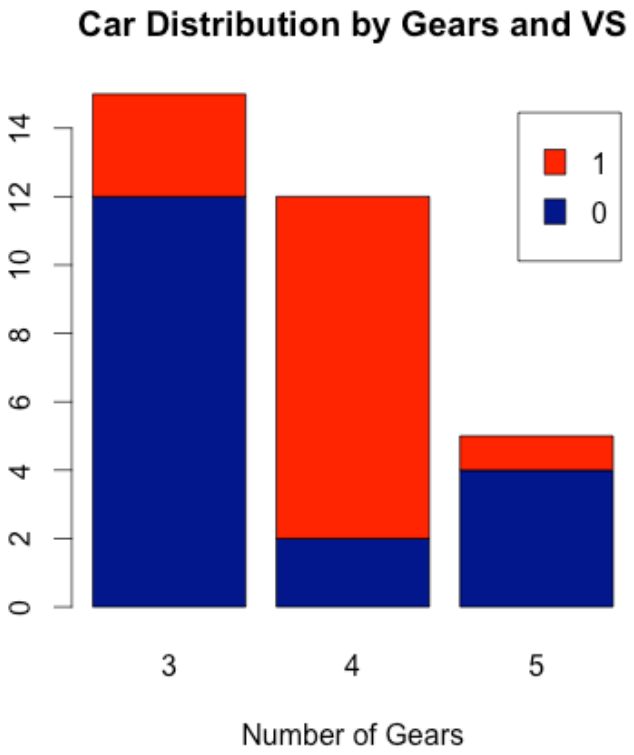
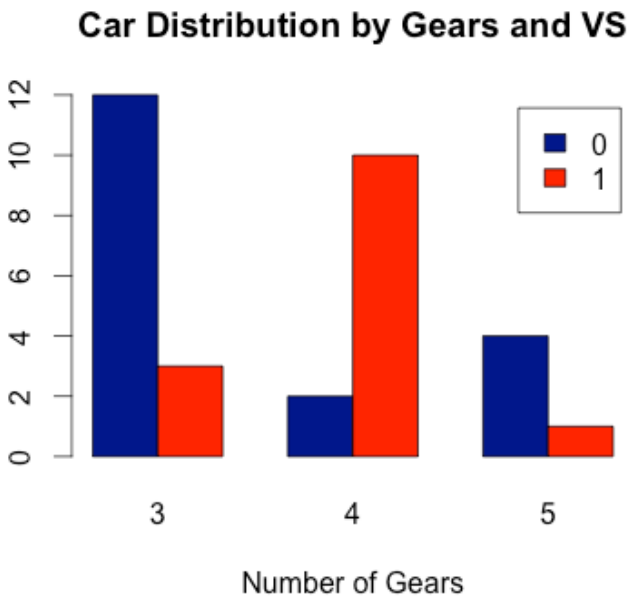
Barplot

- `counts <- table(mtcars$gear)`
- `barplot(counts, main="Car Distribution", xlab="Number of Gears")`



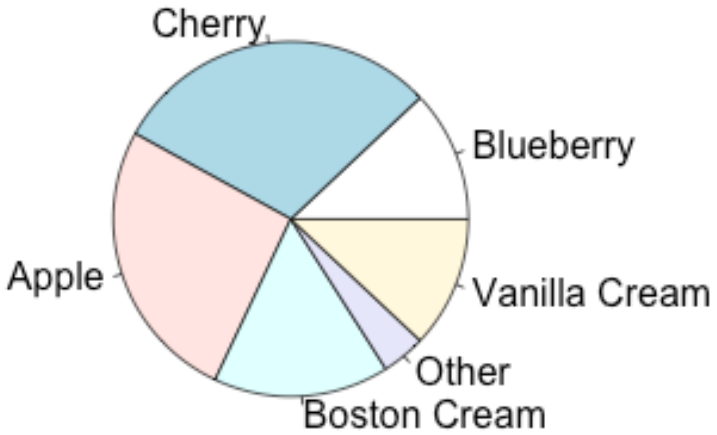
Barplot

- ```
  barplot(counts,
  main="Car Distribution",
  horiz=TRUE,
  names.arg=c("3 Gears",
  "4 Gears", "5 Gears"))
  ```


  ```

Barplot

- `counts <- table(mtcars$vs, mtcars$gear)`
- `barplot(counts, main="Car Distribution by Gears and VS", xlab="Number of Gears", col=c("darkblue","red"), legend = rownames(counts))`



Barplot

- ```
  barplot(counts,
  main="Car Distribution
  by Gears and VS",
  xlab="Number of Gears",
  col=c("darkblue","red"),
  legend =
  rownames(counts),
  beside=TRUE)
  ```


  ```

Pie

- `pie.sales <- c(0.12, 0.3, 0.26, 0.16, 0.04, 0.12)`
- `names(pie.sales) <- c("Blueberry", "Cherry", "Apple", "Boston Cream", "Other", "Vanilla Cream")`
- `pie(pie.sales)`



Pie

- `pie(pie.sales, col = rainbow(6))`

Export plots

- PDF
- PNG
- SVG
- JPEG
- BMP
- TIFF

- `pdf("archivo.pdf")`
- *…R sentences…*
- `dev.off()`

- `?png`
- `?pdf`

Bibliography

- R graphics. Paul Murrell. Computer Science and Data Analysis Series. Chapman & Hall/CRC. 2006.
- http://www.r-bloggers.com
- http://www.statmethods.net

EXERCISES (VISUALIZATION)

Introducción a la Ciencia de Datos

Exercise 1

- Given this data:

stretch	distance
46	148
54	182
48	173
50	166
44	109
42	141
52	166

① Plot distance against stretch.

Exercise 2

The table on the right have ten observations, taken during the years 1970-79, are on October snow cover for Eurasia (snow cover is in millions of square kilometers).

① Plot snow.cover versus year.

② Plot a histogram of the snow.cover values.

year	snow.cover
1970	6.5
1971	12.0
1972	14.9
1973	10.0
1974	10.7
1975	7.9
1976	21.9
1977	12.5
1978	14.5
1979	9.2

Exercise 3

Given the data in NY.xls

① (Optional) convert ºF to ºC and in to mm.

② Plot Year vs. Warmest Minimum Temperature

③ Plot Year vs. Warmest Minimum Temperature and Coldest Minimum Temperature. Don't forget to add a legend!

Year	Lowest Temp	Highest Temp	Warmest Min	Coldest Mini	Average Min	Average Max	Mean Temp	Total Precipi	Total Snowfa	Max 24Hr Pre	Max 24Hr Sno
2014	4	58	44	18	21.8	35.2	28.5	2.34	15.2	0.5	11
2013	11	61	43	20	29.3	40.8	35.1	2.76	1.5	0.9	1.5
2012	13	62	46	27	30.2	44.2	37.2	3.23	4.3	1.38	4.3
2011	6	53	40	24	24.6	34.7	29.7	4.93	36	1.29	12.3
2010	13	57	49	20	27	38.1	32.3	2.08	2.1	1.25	1.3
2009	6	47	38	16	22.4	33.5	27.9	2.98	9	1.19	3
2008	12	66	53	20	30.6	42.4	36.5	2.85	0	1.29	0
2007	9	72	54	22	31.4	43.6	37.5	3.63	2.6	1.52	1
2006	16	66	45	32	36.2	47.5	41.9	4.93	3	1.17	2
2005	5	66	49	16	25.2	37.4	31.3	4.67	15.3	1.03	8.5
2004	1	58	44	15	18.5	31	24.7	2.13	17.3	0.56	6
2003	7	50	38	20	22.1	32.9	27.5	2.3	4.7	1.52	1
2002	21	69	50	31	34.5	45.4	39.9	1.93	3.5	0.62	3
2001	18	54	40	27	28.5	38.8	33.6	3.16	8.3	0.83	6
2000	3	64	51	19	24.7	37.9	31.3	3.23	9.5	0.84	5.5
1999	9	60	41	22	27	40.8	33.9	7.02	4.5	2.42	2.5
1998	14	65	47	29	34.2	45.8	40	5.2	0.5	2.55	0.5
1997	6	62	44	17	25.1	39.2	32.1	3.65	4.4	1.23	3.5
1996	6	56	40	18	24	37	30.5	5.94	26.1	1.4	13.6
1995	15	64	54	27	31.9	43.1	37.5	3.75	0.2	1.41	0.2
1994	-2	55	35	10	17.9	33.2	25.5	5.62	12	1.87	4.5
1993	21	64	41	27	30.1	42.5	36.3	3.44	1.3	1.42	1.1
1992	11	62	42	23	28.9	42.5	35.7	1.68	1.5	0.5	1.3
1991	10	55	43	20	28.5	41.2	34.9	3.38	8.4	0.85	5.7
1990	24	66	46	33	35.2	47.5	41.4	5.34	0.9	1.11	0.8
1989	11	61	44	26	30.7	44	37.4	2.29	5	0.6	5
1988	5	53	41	17	23.2	35.8	29.5	3.64	13.9	0.8	5.2
1987	8	57	46	19	27.1	37.5	32.3	5.81	13.6	1.68	8.1
1986	8	69	45	22	27.4	40.8	34.1	4.23	2.2	2.19	1.4
1985	-2	58	40	9	23	34.6	28.8	1	8.4	0.3	4.1
1984	8	57	39	19	24.6	35.1	29.9	1.87	11.7	0.6	4.2
1983	12	58	43	21	29.4	39.6	34.5	NR	1.9	NR	1
1982	0	58	37	15	19.6	32.5	26.1	6.46	11.8	2.73	5.8
1981	2	49	40	14	20.2	32.3	26.2	0.58	8	0.41	4.1
1980	17	60	40	27	28	39.4	33.7	1.72	2	0.81	2
1979	8	63	50	20	26.9	40.2	33.5	10.52	6.6	3.45	3.1
1978	12	58	36	21	22	33.9	28	8.27	34	1.8	13.7
1977	-2	44	28	12	16.4	27.7	22	2.25	13	1.24	5
1976	-1	56	36	15	20.5	34.2	27.3	5.78	5.6	1.94	1.9
1975	15	63	48	29	31.3	43.3	37.3	4.76	2	0.8	1.3
1974	6	66	45	23	29.1	41.4	35.2	3.8	13.4	1.05	5.6
1973	10	66	45	18	28.8	42.1	35.5	4.53	1.8	1.49	1.8
1972	3	62	45	15	28	42.1	35	2.41	2.8	0.84	1.8
1971	4	66	34	16	20.8	33.1	26.9	2.67	11.4	0.99	6.6
1970	3	51	42	14	18.7	31.5	25.1	0.66	6.4	0.16	2.6
1969	11	48	42	24	26.3	37.3	31.8	1.1	1	0.43	1
1968	-1	50	36	13	20.1	33.3	26.7	2.04	3.6	1.3	1.5
1967	13	66	54	25	31.5	43.2	37.3	1.39	1.4	0.82	0.5
1966	8	62	52	19	25.7	38.8	32.2	2.63	11.6	0.72	5.8
1965	9	55	41	16	23.7	35.6	29.6	3.09	14.8	0.64	6.3
1964	9	55	41	22	29.5	41.8	35.6	4.92	13.3	1.29	11.5
1963	6	51	39	20	24	36.2	30.1	1.93	5.3	0.38	3
1962	11	56	41	25	25.5	39.6	32.6	2.62	0.6	1.57	0.4
1961	5	55	42	19	21.9	33.5	27.7	1.88	16.7	0.78	5.5
1960	17	54	38	28	28.8	38.9	33.9	2.4	2.5	0.84	1.8
1959	14	61	38	22	24.3	38	31.2	2.34	1.5	0.83	0.9
1958	12	53	39	22	26.3	37.5	31.9	3.79	9.2	0.86	4.5
1957	0	60	43	12	22.2	34.8	28.5	1.7	8.9	0.49	2.9
1956	14	48	43	28	26.7	37.3	32	1.54	1.2	0.52	0.5
1955	13	50	40	24	25.2	36.7	30.9	0.77	2.6	0.31	0.9
1954	7	56	41	21	23.5	38.1	30.8	1.85	12.7	0.45	5.4
1953	21	57	49	30	31.8	43.4	37.6	4.9	4.1	1.3	2
1952	8	60	45	23	28.9	43.4	36.2	4.58	6.1	0.85	5.8
1951	11	66	47	27	29.4	44.5	36.9	3.31	0.9	1.27	0.9
1950	16	72	59	31	36.2	48.6	41.4	2.19	0.4	0.46	0.4
1949	20	61	43	30	32.6	44.6	38.4	5.98	6.8	1.7	5.5
1948	0	45	33	16	19.3	31.2	25.2	5.97	15.3	1.23	3.5
1947	12	63	42	23	30.2	44.1	37.1	2.74	5.5	0.93	4
1946	5	62	54	26	27	41.2	34.1	1.86	4.2	0.69	1.8
1945	2	55	33	15	17.9	32.5	25.2	2.24	12.3	0.77	6.7
1944	12	62	38	23	27.8	40	33.9	3.2	4.8	0.99	3.2
1943	10	56	38	25	24.3	37.2	30.7	2.46	9.5	1.25	7.1
1942	5	57	43	14	23.9	37.5	30.7	2.88	6.4	1.09	2.8
1941	10	47	36	25	23.2	35.6	29.4	3.23	9.2	1.18	5
1940	7	54	36	19	18.9	31	25	2.81	3.5	2.06	1.5
1939	6	55	44	18	26.5	37.8	32.1	3.8	10.3	1.19	7.8
1938	6	57	42	17	25.9	38.3	32.1	3.99	6.5	1.35	2.1
1937	25	66	50	32	33.7	46.9	40.3	5.97	6.5	0.97	5.7
1936	-3	54	37	16	23.3	35.3	29.3	6.9	12.1	2.39	9
1935	-1	58	47	16	21.6	36.1	28.8	3.91	23.6	1.09	12.8
1934	6	58	41	16	28.5	40.8	34.8	3.42	0.1	1.24	0.1
1933	13	61	48	28	34	46.6	40.4	1.78	0	0.82	0
1932	24	70	56	33	37.7	48.6	43.2	4.85	0.8	1.33	0.6
1931	11	56	41	24	26.3	39.2	32.7	2.56	0.5	0.93	0.3
1930	10	64	52	22	27	39.5	33.3	2.37	3.5	0.81	1.8
1929	8	62	38	24	25	38.8	31.9	3.33	2.3	0.78	0.9
1928	12	58	44	24	26.9	41	34	1.87	2.7	0.64	1.8
1927	-1	52	40	22	24	37.8	30.9	2.03	5.7	0.44	5.1
1926	5	54	40	23	25.7	38.2	32	2.85	3.1	0.95	2.3
1925	-2	45	33	16	22	34.8	28.4	5.42	27.4	1.92	11.5
1924	5	57	40	18	24.9	40.4	32.6	4.9	2.5	1.44	2.4
1923	9	53	44	21	23.6	36.3	29.9	6.99	24.5	1.68	7.8
1922	7	50	37	16	21.9	34.4	28.2	2.73	9.4	1.38	3.4
1921	4	57	42	16	25.7	40.8	33.2	2.74	3.5	1.75	1.8
1920	-1	51	36	14	16.1	30.8	23.4	2.09	8.2	0.64	2.8
1919	10	61	43	26	29.4	41.8	35.6	3.51	0.3	1.22	0.3
1918	-4	53	29	10	15.6	27.7	21.7	3.53	13.2	1.15	5.9
1917	8	51	40	20	26.1	39.6	32.8	2.96	5.8	0.57	1.8
1916	7	69	53	21	28.3	43.6	36	1.28	3.4	0.3	3
1915	12	58	47	24	28.7	40.2	34.5	7.94	5.8	2.35	1.6
1914	-5	58	42	9	24.8	38.2	31.5	5.27	1.3	1.51	1.1
1913	18	65	45	29	33.5	48.3	40.8	1.43	0.3	0.54	0.2
1912	3	54	38	8	16.9	30.5	23.7	2.56	13	0.93	3.5
1911	18	56	49	26	30.6	41.9	36.2	2.79	1.3	0.76	1
1910	5	48	40	27	25.7	37.4	31.5	4.64	2.1	1.18	1
1909	5	57	49	24	28.1	40.8	34.5	3.32	11.4	1.5	4.3
1908	8	50	41	18	26.2	39.4	32.9	3.76	10	1.46	10
1907	4	64	49	20	28	41.8	34.9	2.89	11	0.96	6
1906	14	62	50	29	33.3	43.5	38.4	2.69	1.5	0.85	1.5
1905	4	53	42	16	23.6	34.9	29.2	2.77	18.4	0.91	11
1904	-4	54	39	13	15.5	31.1	25.3	2.97	11.1	0.72	7.5
1903	10	51	39	21	24.3	36.7	30.5	4.18	4.5	1.3	4
1902	13	51	40	21	24.7	36.3	30.8	2.27	6.1	0.8	3
1901	8	53	38	27	26.8	38.2	32.6	1.99	2	0.9	1.8
1900	10	57	46	20	25.5	38.8	32.4	4.22	1	1.46	0.5