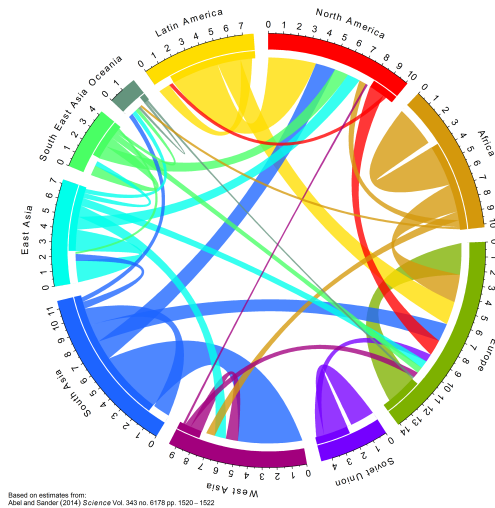


# VISUALIZATION

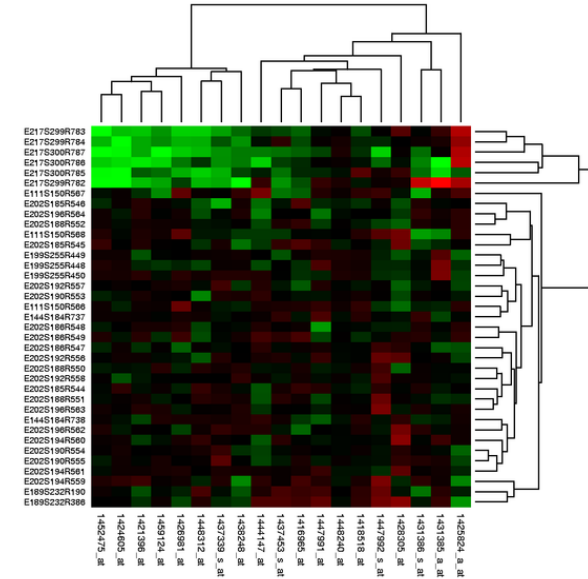
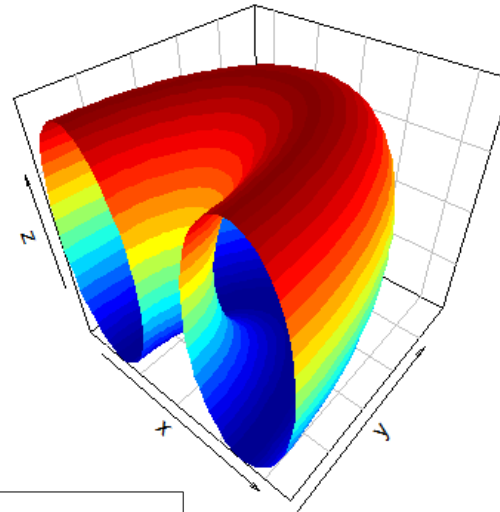
---

Introducción a la Ciencia de Datos

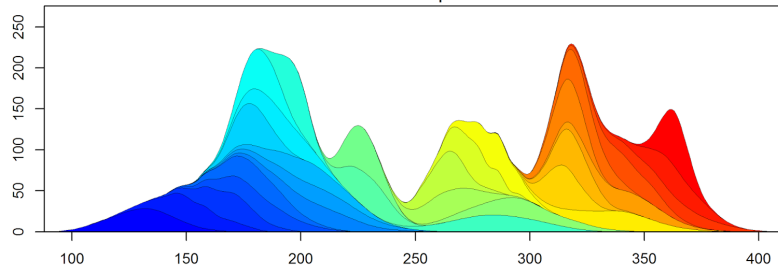
# Gallery of plots



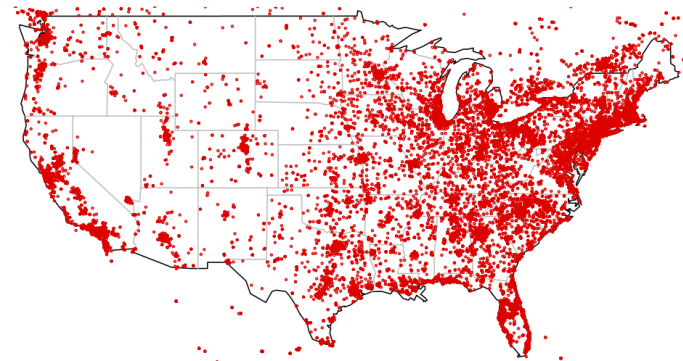
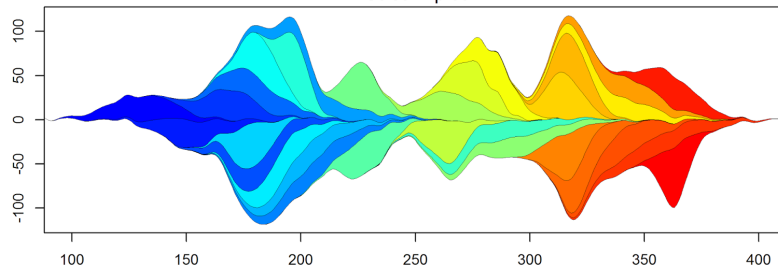
Half of a Torus



Stacked plot



Stream plot



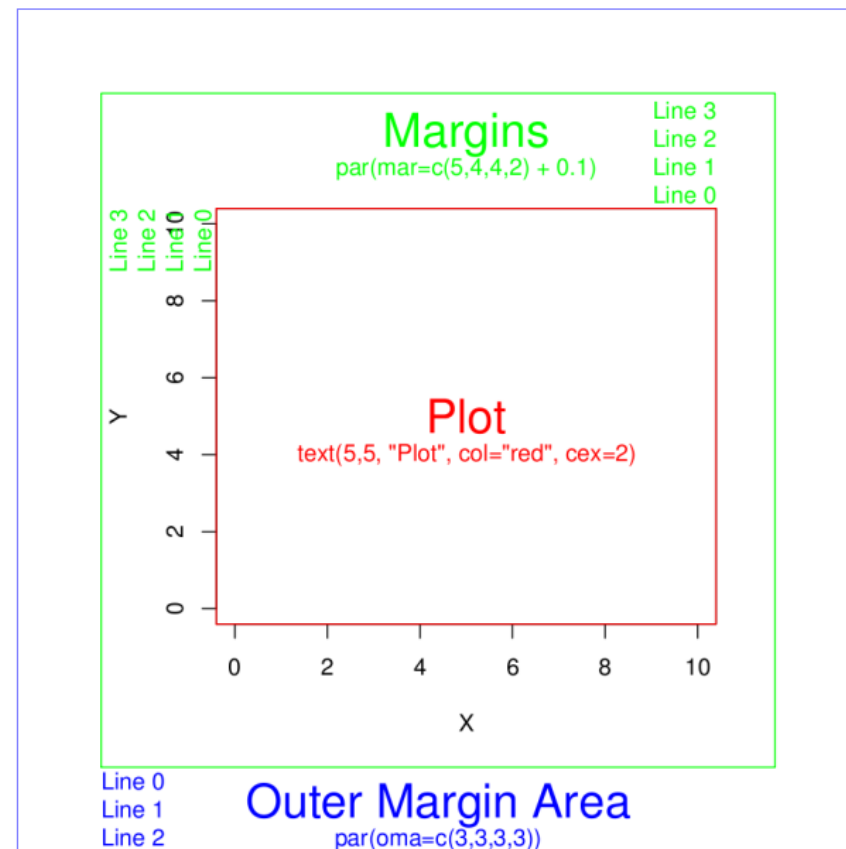
# R plots

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

# Plot

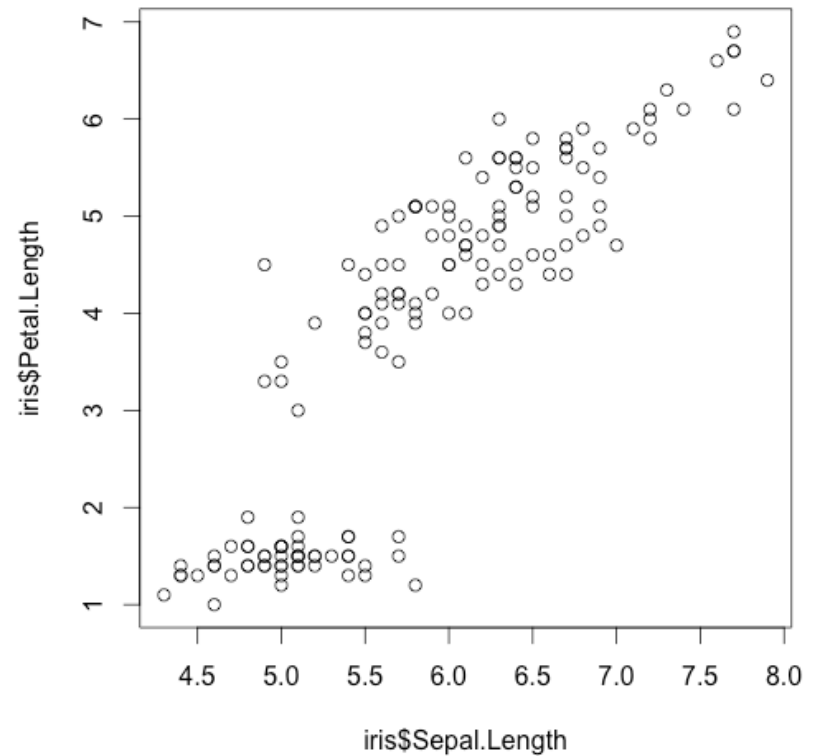
- To change parameters for margins and plot areas use the `par` function.

- [Link to R manual](#)



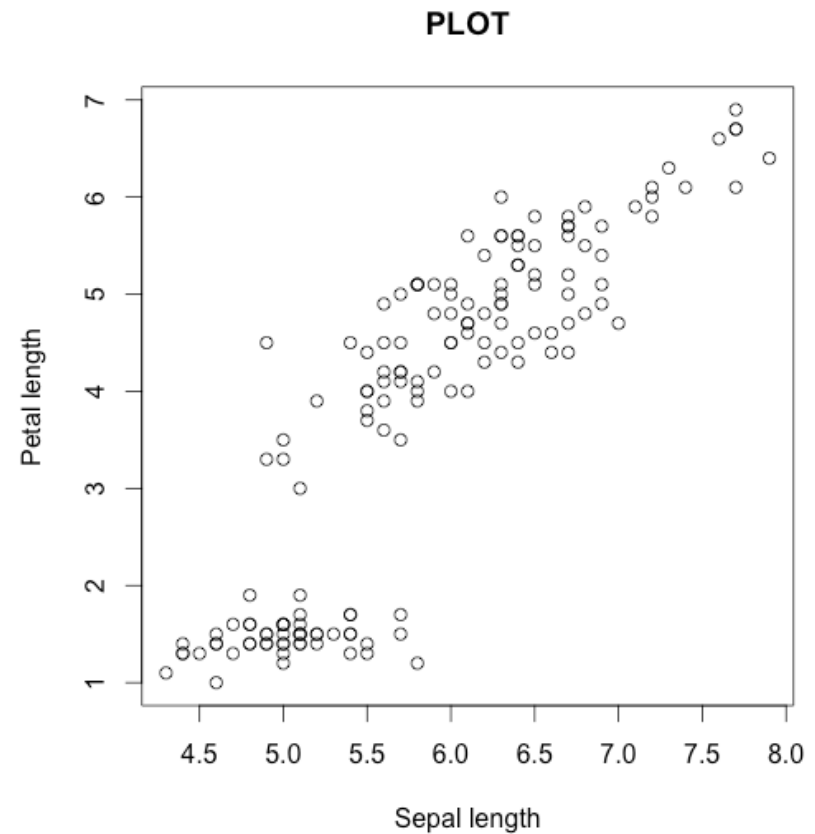
# Plot

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



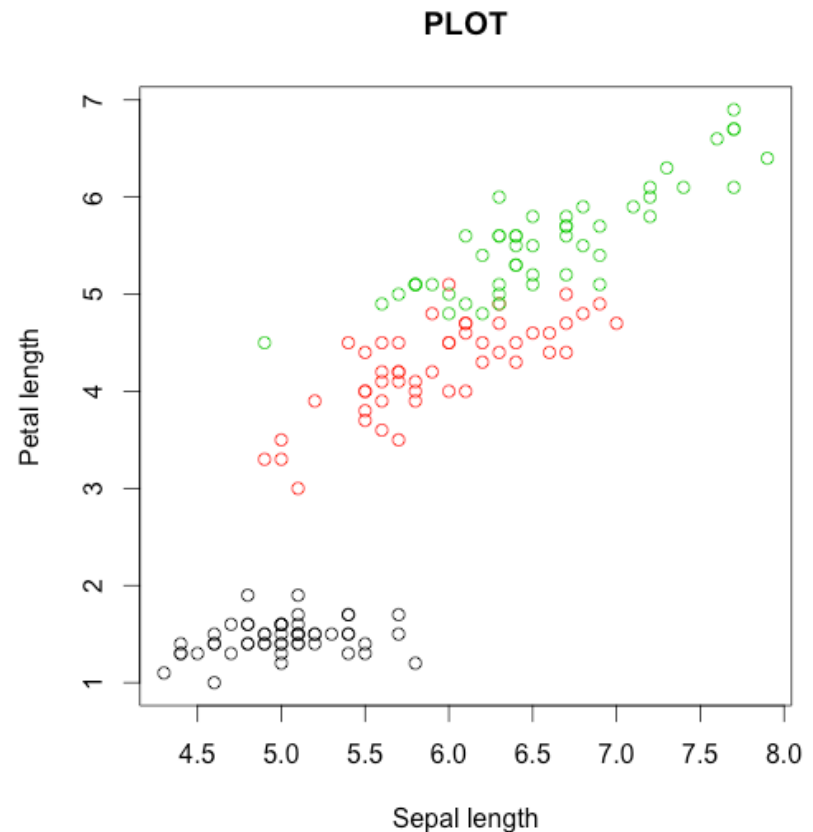
# Plot

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



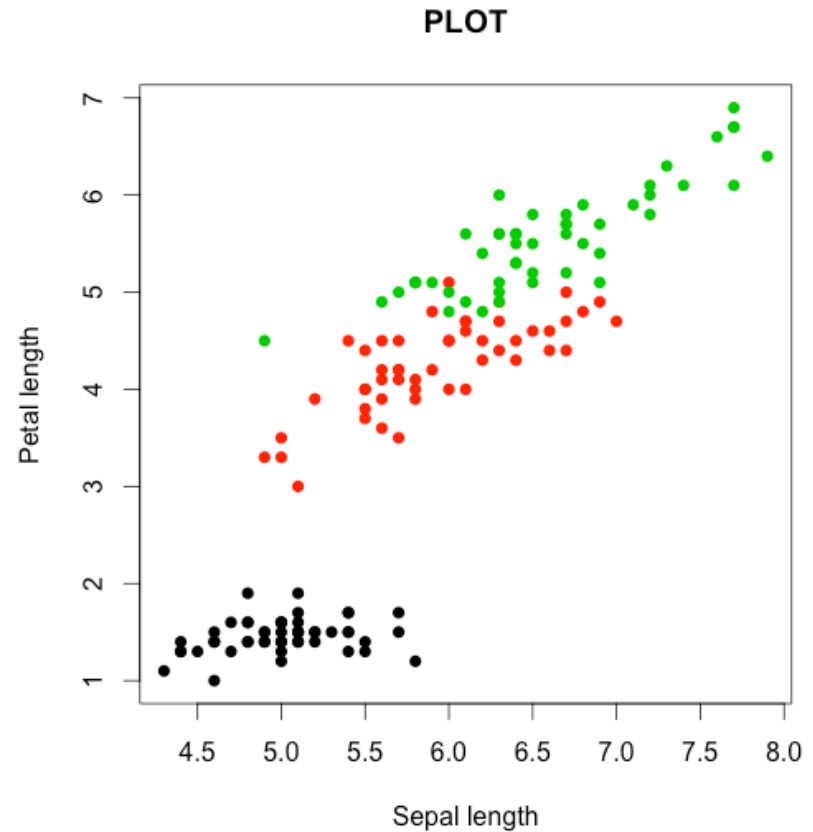
# Plot

- `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")`





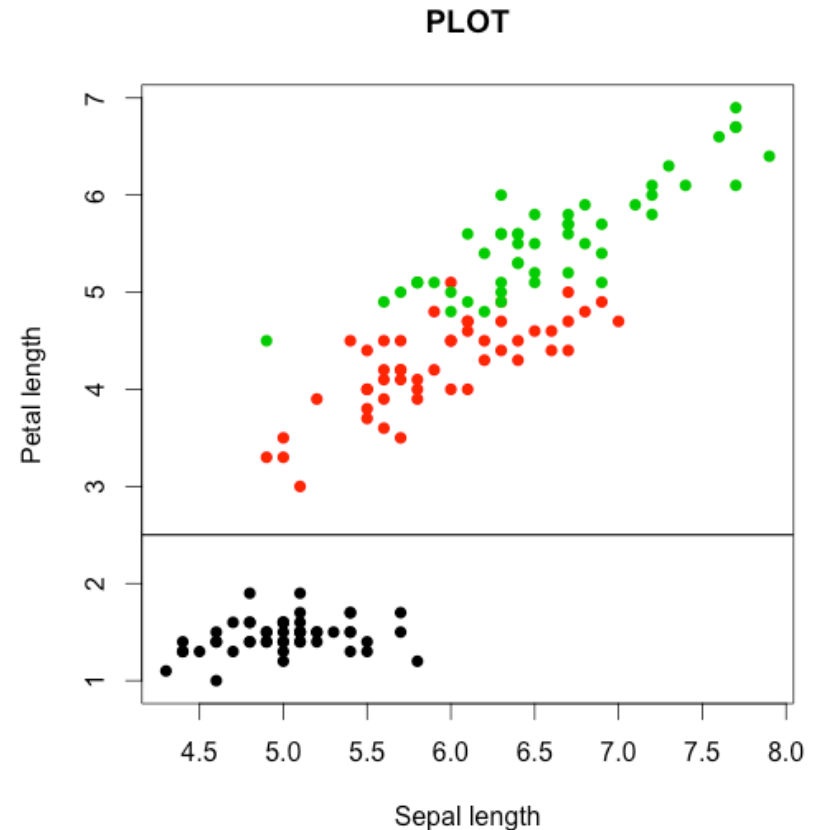
# Plot

pch

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

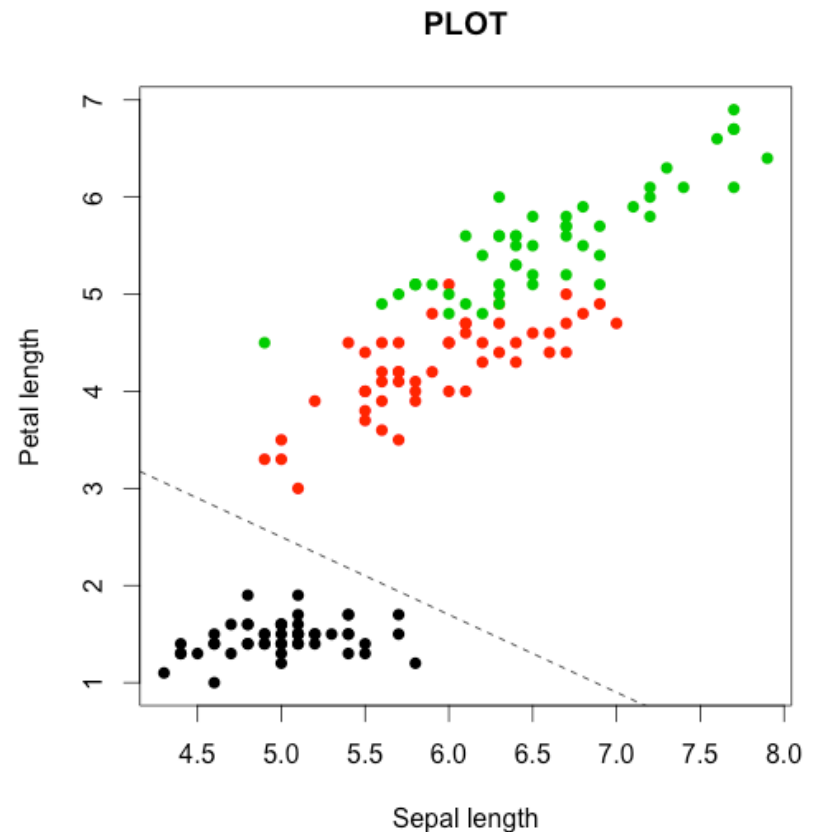
# Plot

- `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

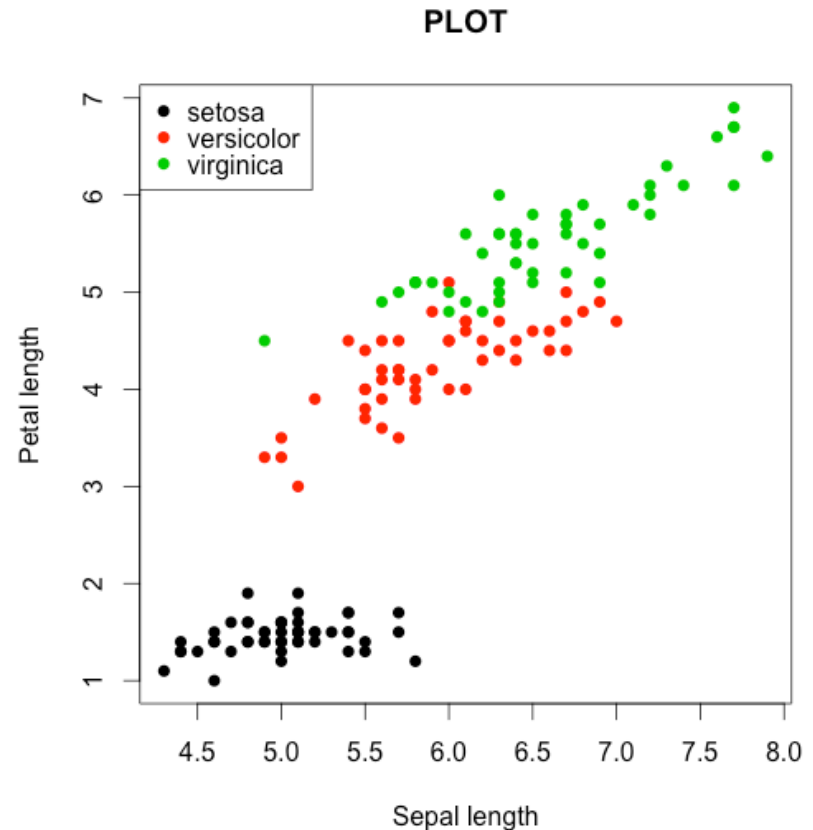
0: 	10: 	20: 	A: 
1: 	11: 	21: 	a: 
2: 	12: 	22: 	B: 
3: 	13: 	23: 	b: 
4: 	14: 	24: 	S: 
5: 	15: 	25: 	`: 
6: 	16: 	@: 	.: 
7: 	17: 	+: 	,: 
8: 	18: 	?: 	?: 
9: 	19: 	#: 	*: 

## lty

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

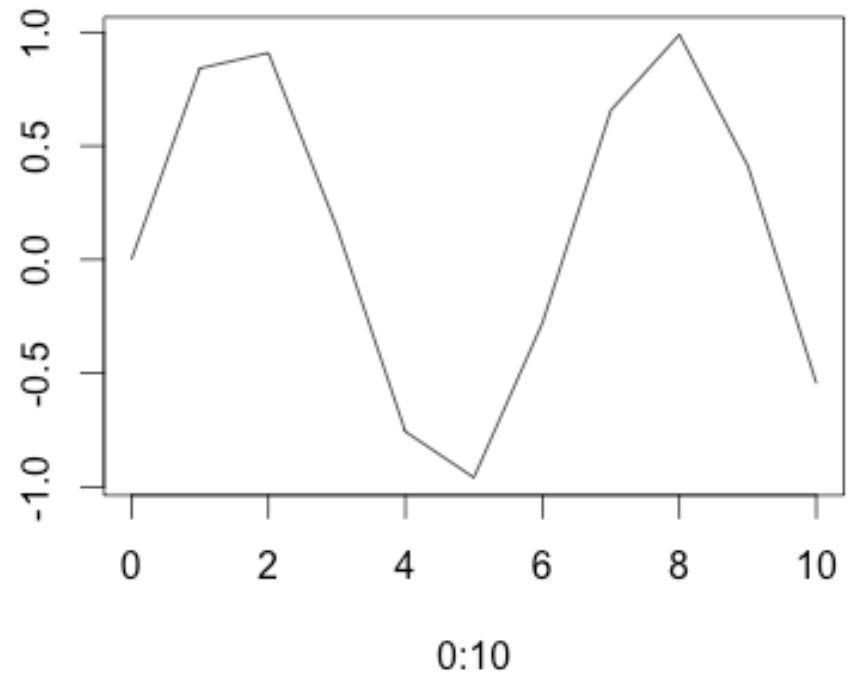
# Plot

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



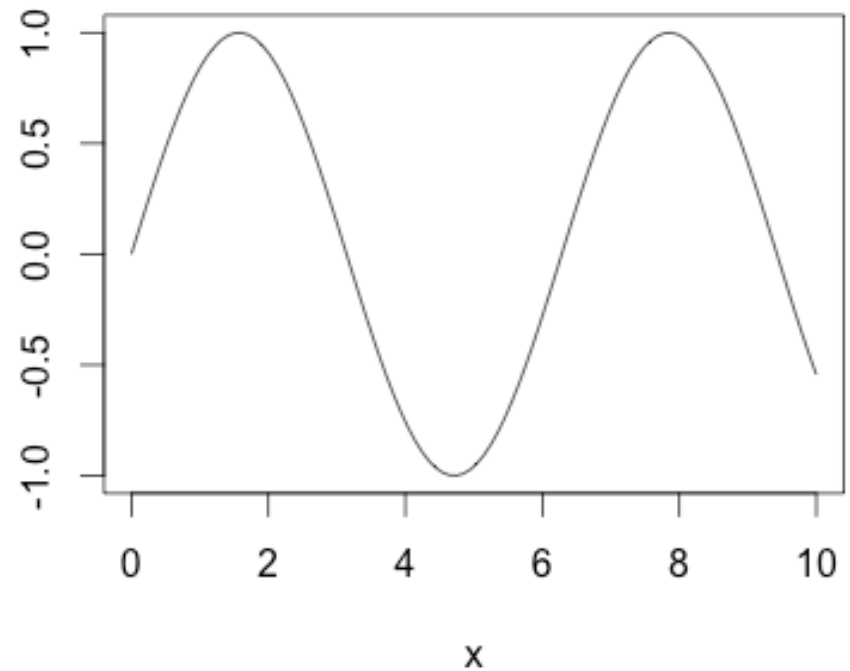
# Curves

- `plot(0:10, sin(0:10), type="l")`



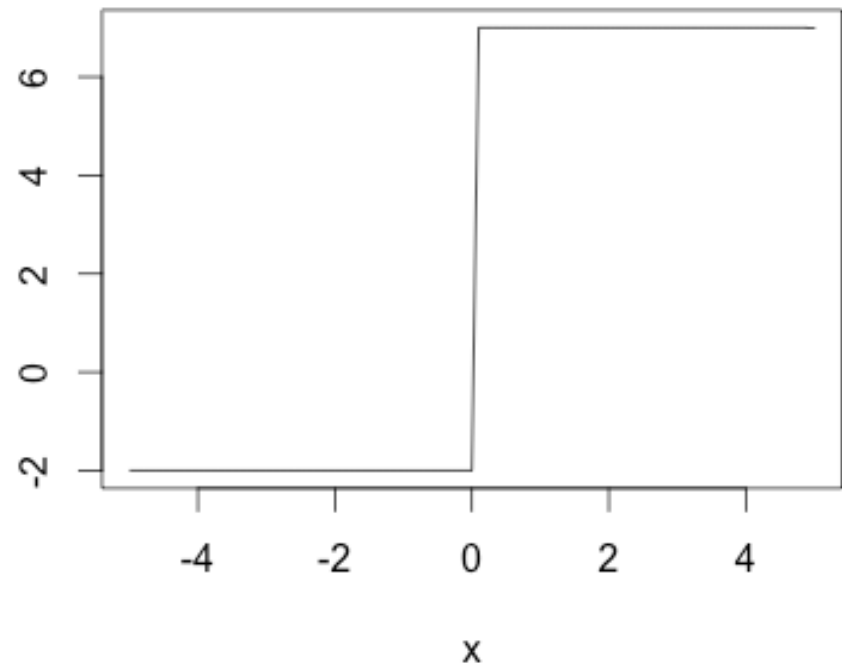
# Curve

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



# Piecewise functions

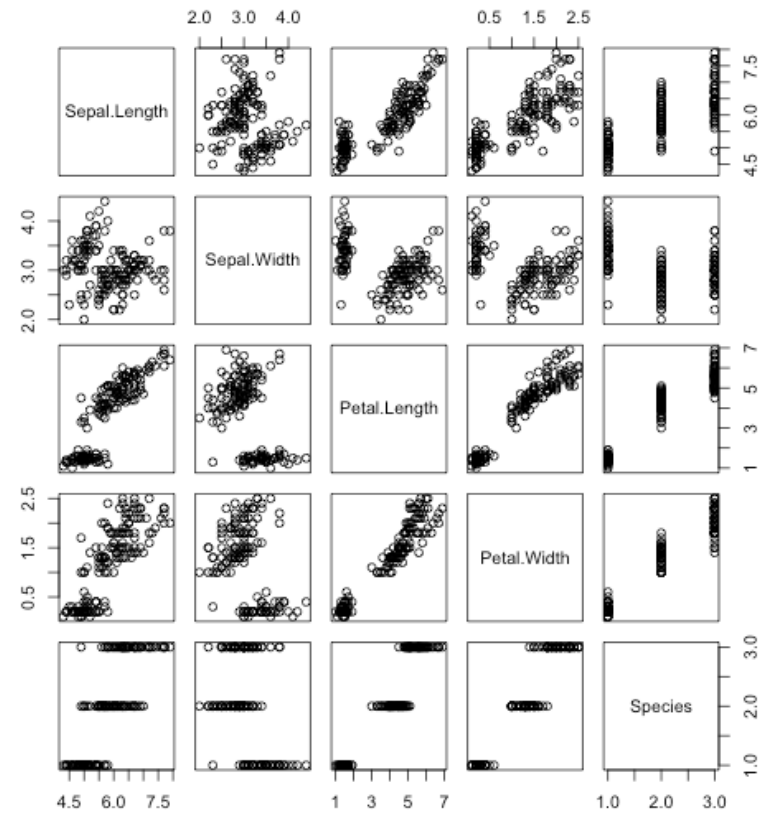
- `myfun <- function(x)`  
  `{ifelse(x > 0, 7, -2)}`
- `curve(myfun(x), -5, 5)`





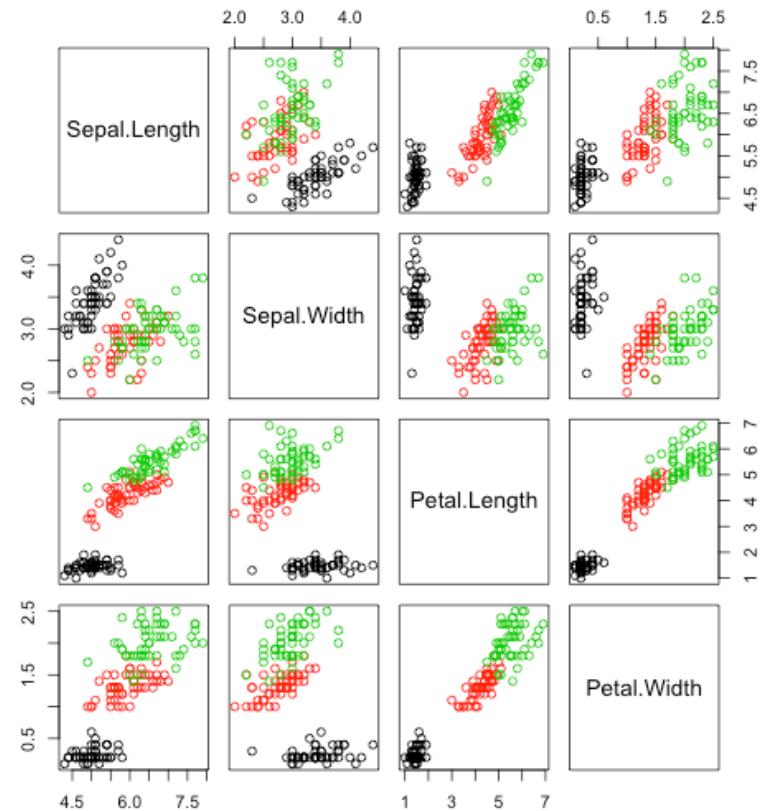
# Scatterplot

- `plot(iris)`



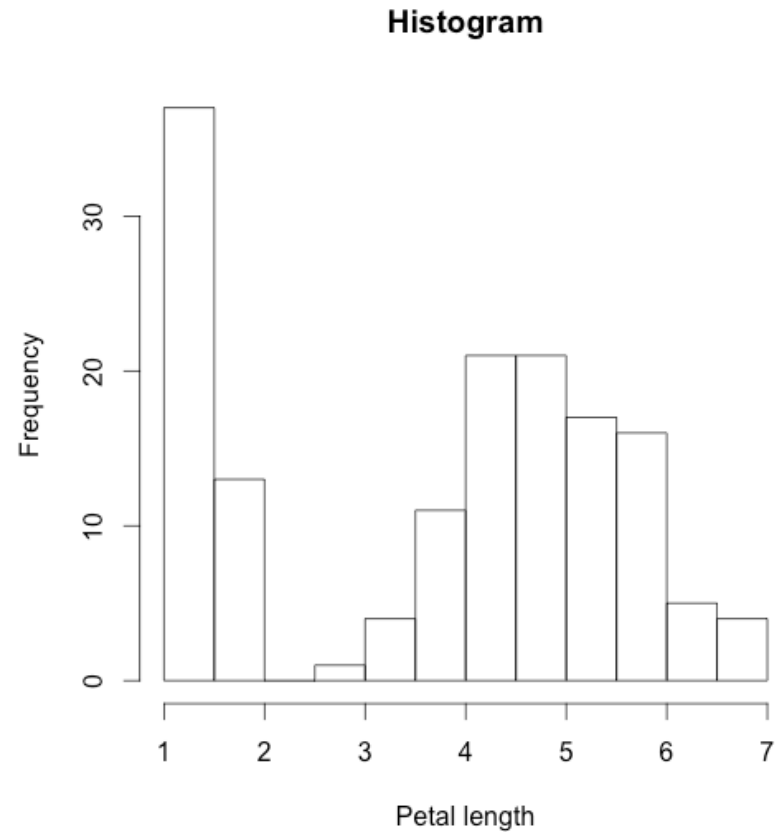
# Scatterplot

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



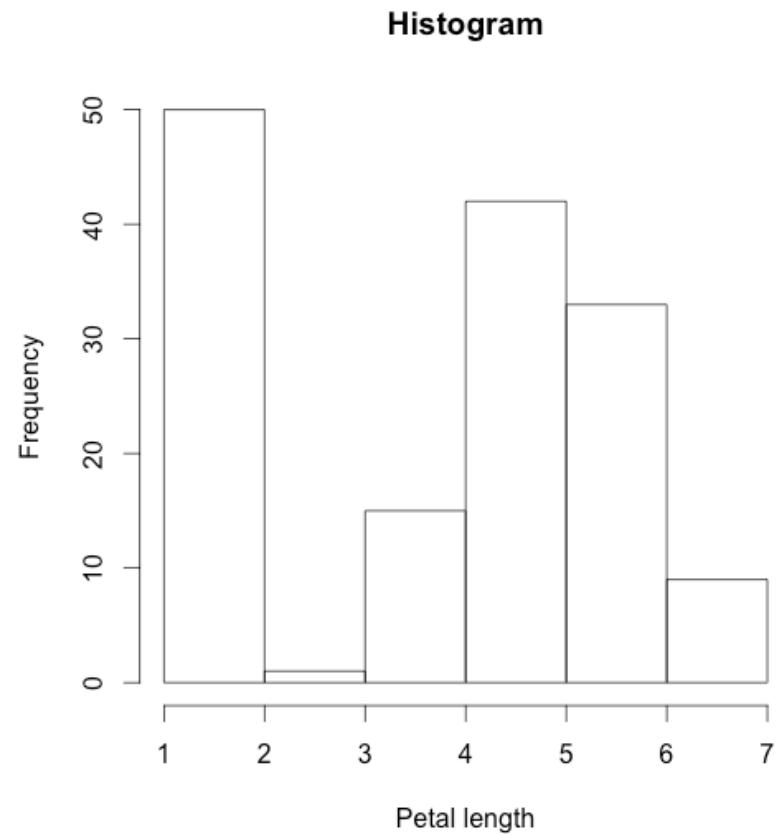
# Histogram

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



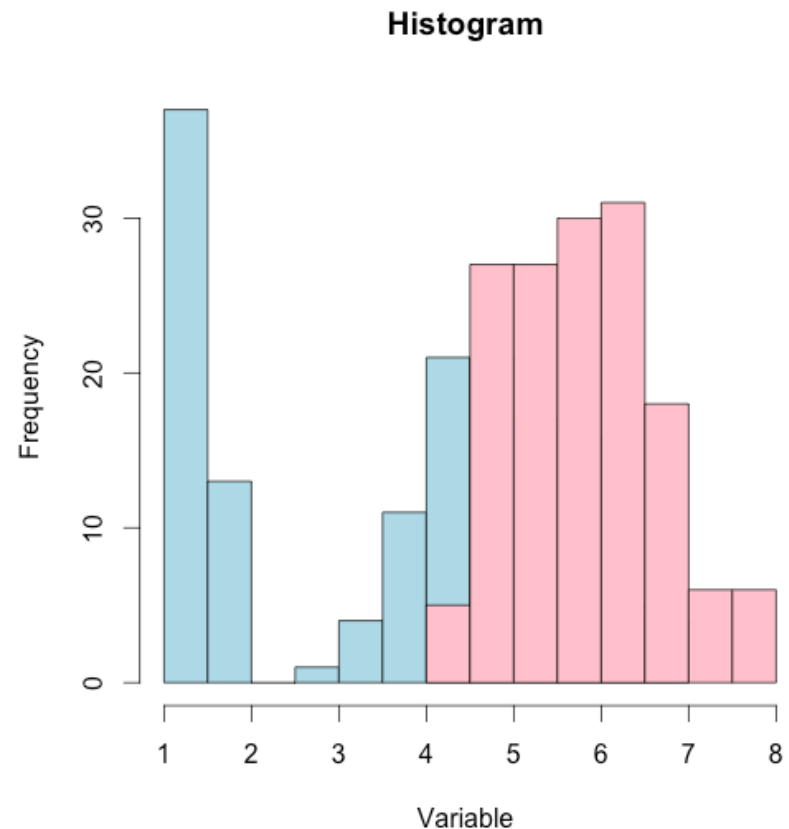
# Histogram

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



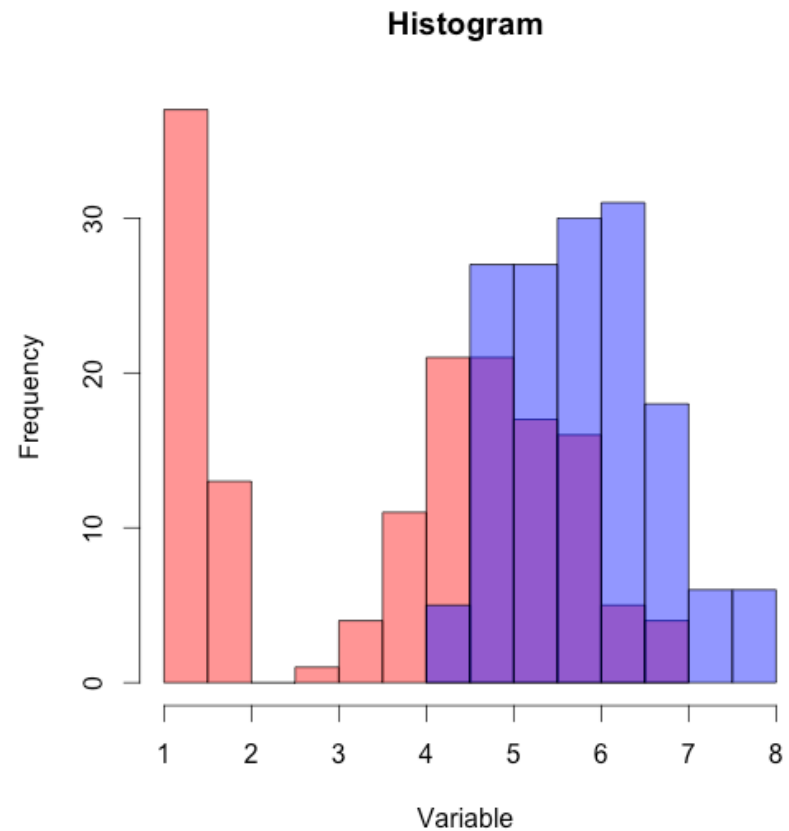
# Histogram

- `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)`

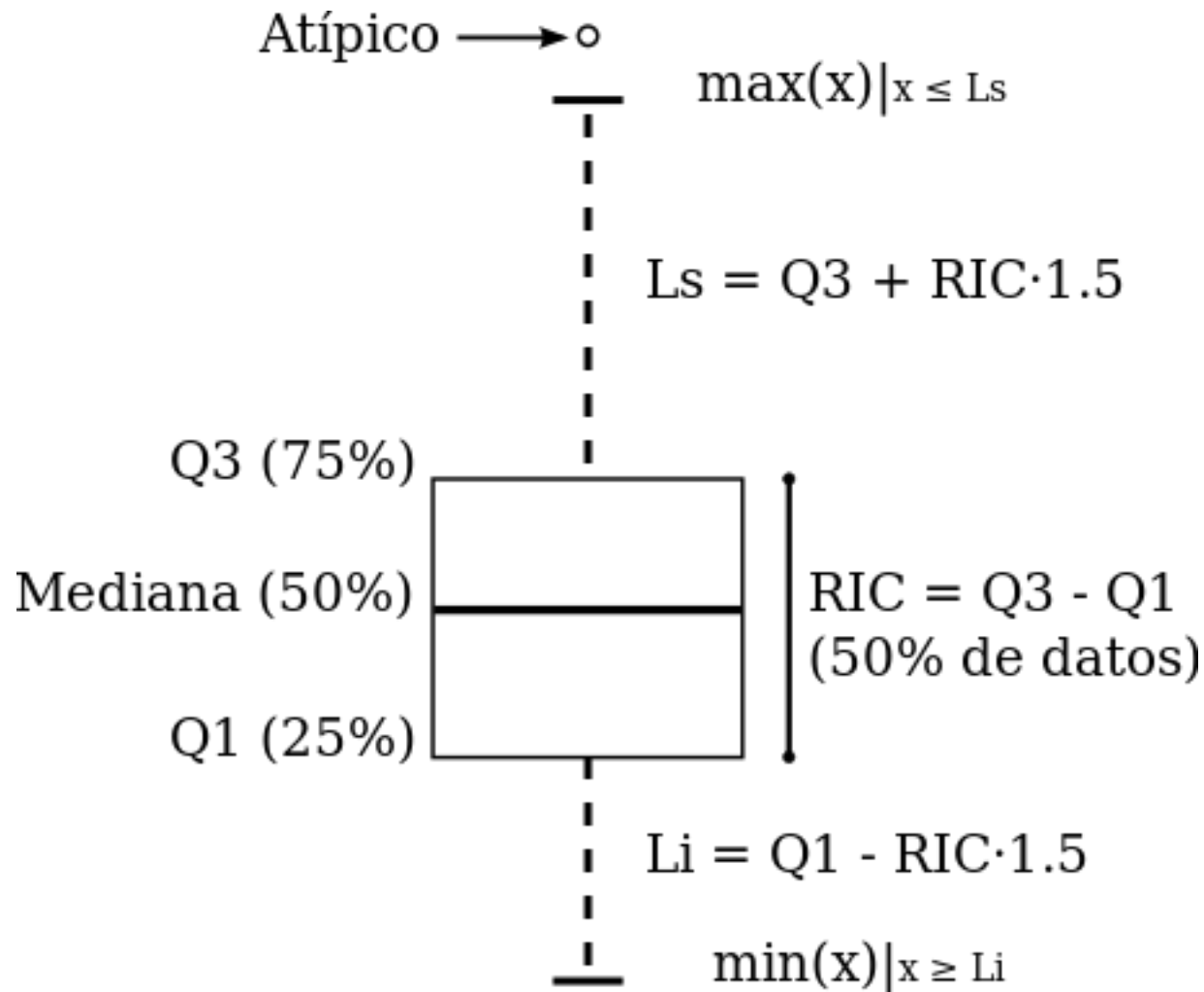


# Histogram

- `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)`

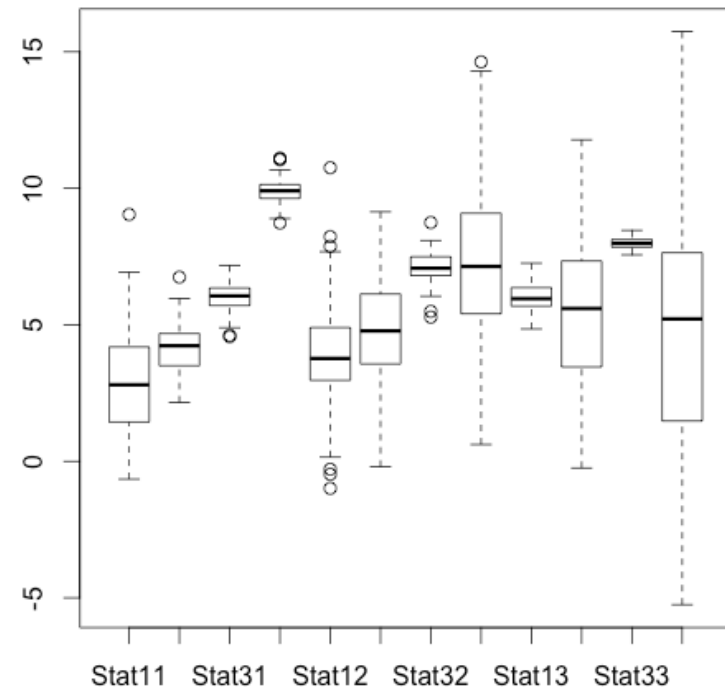


# Boxplot



# Boxplot

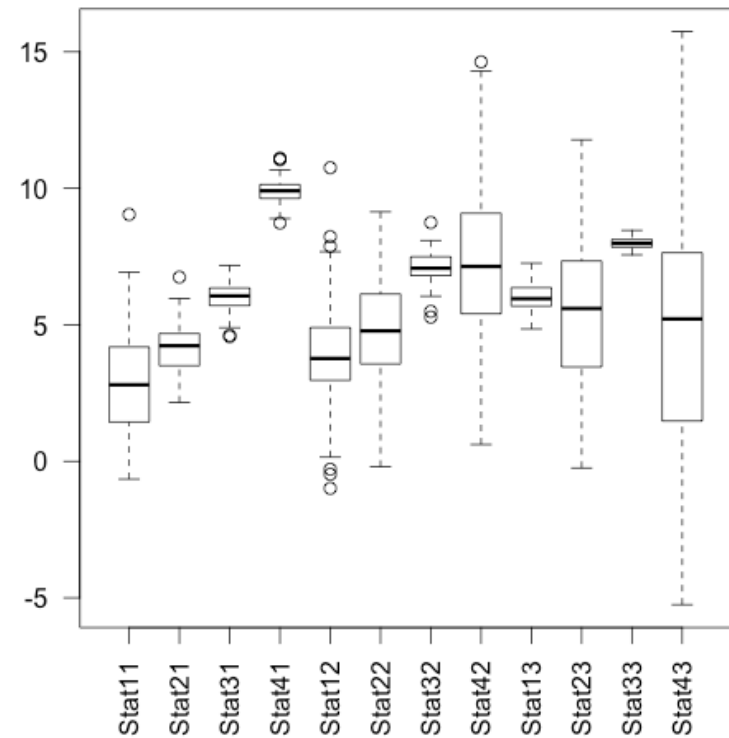
- `data<-`  
`data.frame(Stat11=rnorm(100,mean=3,`  
`sd=2),`  
`Stat21=rnorm(100,mean=4,sd=1),`  
`Stat31=rnorm(100,mean=6,sd=0.5),`  
`Stat41=rnorm(100,mean=10,sd=0.5),`  
`Stat12=rnorm(100,mean=4,sd=2),`  
`Stat22=rnorm(100,mean=4.5,sd=2),`  
`Stat32=rnorm(100,mean=7,sd=0.5),`  
`Stat42=rnorm(100,mean=8,sd=3),`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





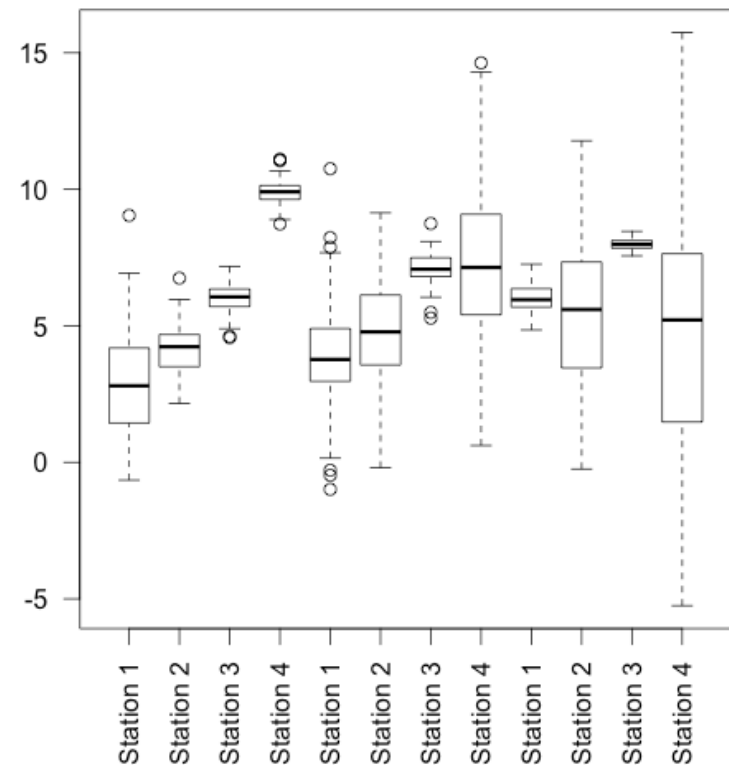
# Boxplot

- `boxplot(data, las = 2)`



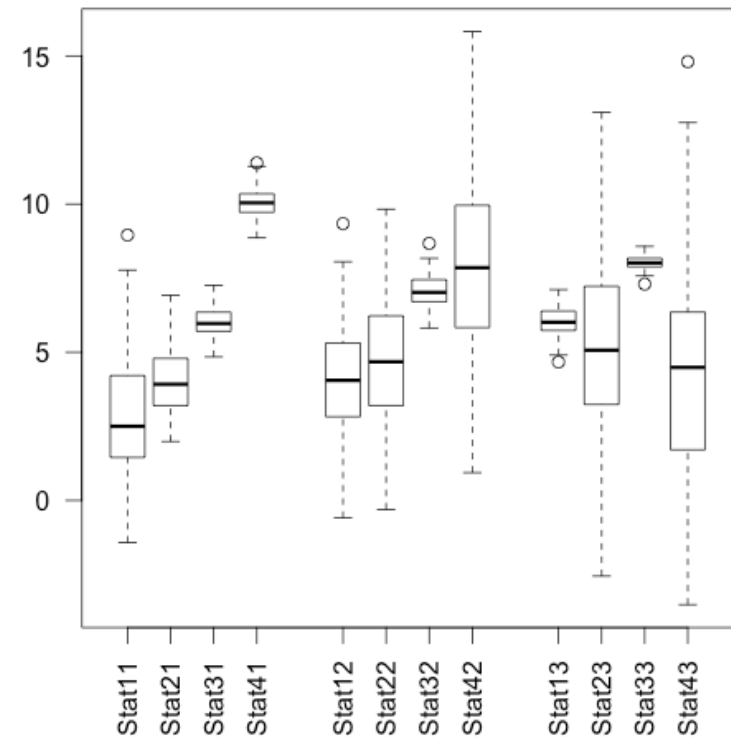
# Boxplot

- `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"))`



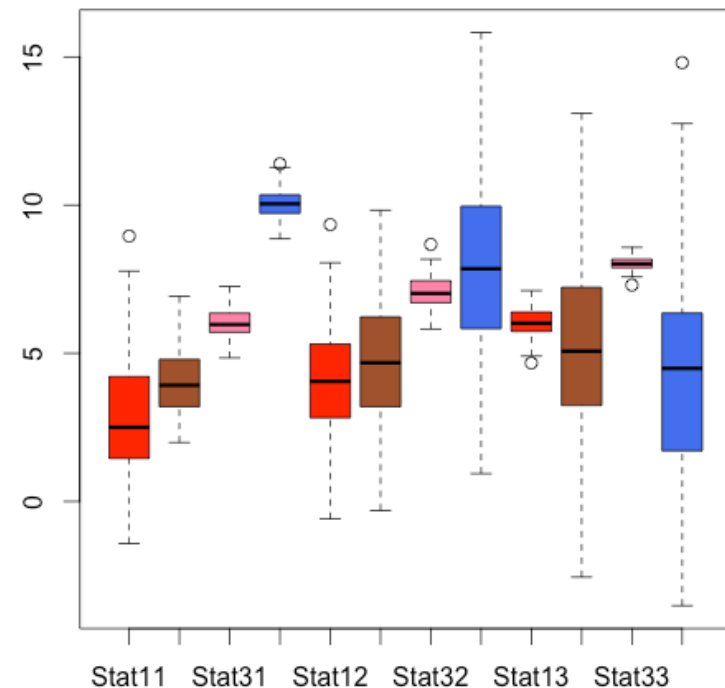
# Boxplot

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



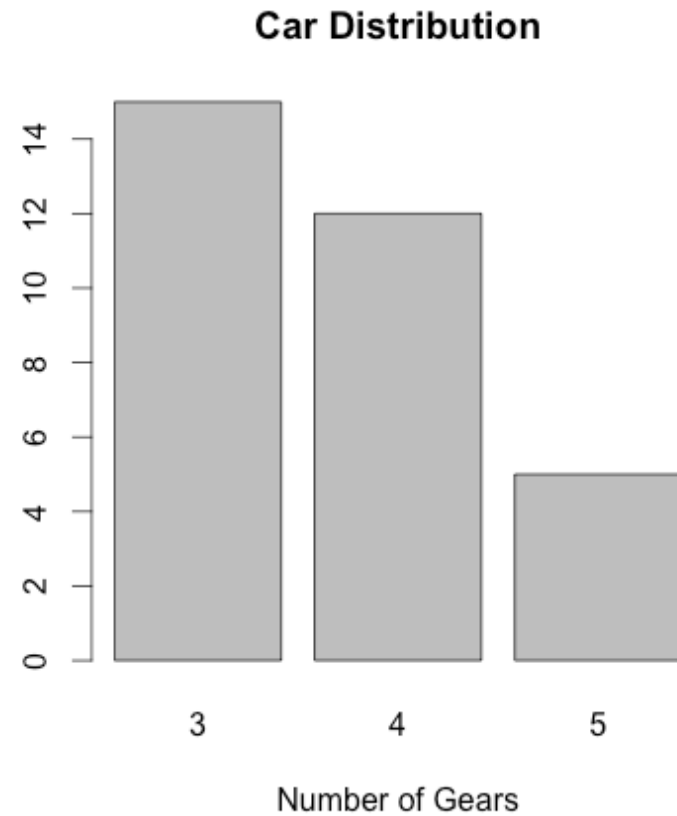
# Boxplot

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



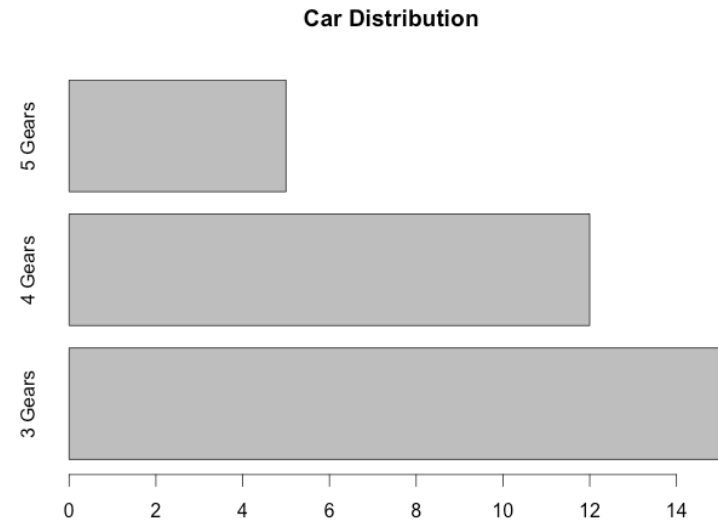
# Barplot

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



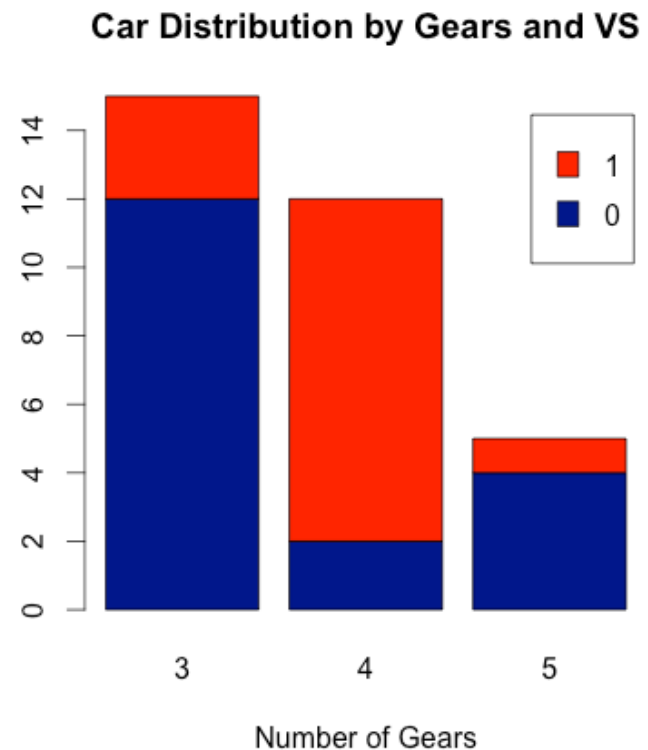
# Barplot

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



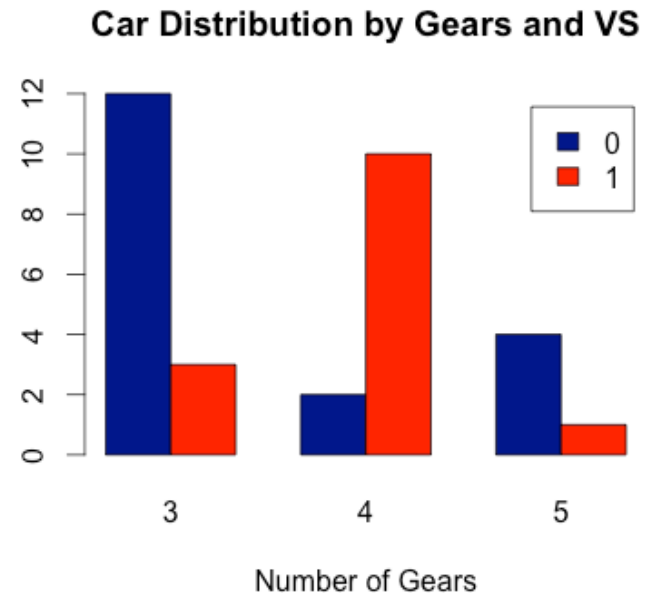
# Barplot

- `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

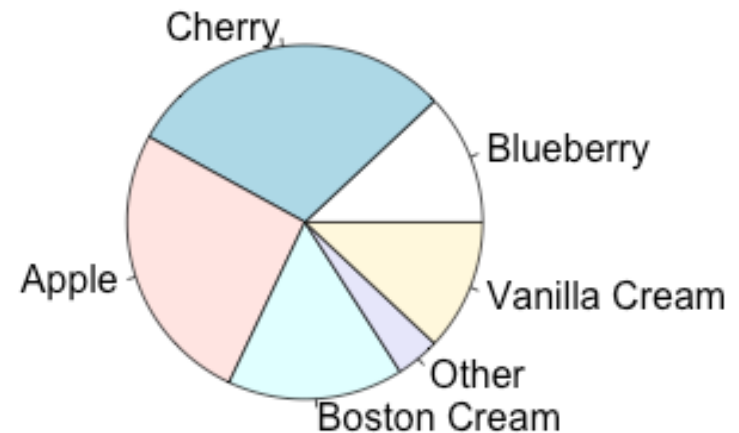
- `barplot(counts,  
main="Car Distribution  
by Gears and VS",  
xlab="Number of Gears",  
col=c("darkblue","red"),  
legend =  
rownames(counts),  
beside=TRUE)`





# 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:

① Plot distance against stretch.

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

## 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 °F to °C 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	Min	Average	Min	Average	Max	Temp	Total	Precip	Total	Snowy	Max	24hr	Precip	24hr	Day
2014	4	16	44	16	21.8	33.2	28.5	2.44	19.2	0.5	1.1	1.1										
2013	11	61	81	20	26.3	40.6	35.1	2.76	1.3	0.8	1.3	1.3										
2012	18	82	86	27	30.2	44.2	37.2	3.23	4.3	1.8	4.3	1.8										
2011	8	53	80	24	26.6	36.7	29.7	4.03	36	1.28	12.3	1.3										
2010	17	49	20	37	36.1	51.5	29.6	2.1	1.28	1.3	1.3	1.3										
2009	6	47	38	10	22.4	33.5	27.9	2.06	9	1.19	1.3	1.3										
2008	12	64	52	20	36.6	42.6	36.5	2.85	0	1.29	0	1.29										
2007	9	72	34	22	32.4	43.6	37.5	3.65	2.6	1.12	1	1.12										
2006	16	84	45	32	34.2	47.5	40.5	4.09	2	1.27	2	1.27										
2005	5	66	49	18	29.2	37.6	31.3	4.07	15.3	1.09	8.5	1.09										
2004	1	18	44	11	28.1	37	24.7	2.12	17.3	0.38	4	0.38										
2003	7	10	38	20	32.1	32.9	27.5	2.3	4.7	1.12	1	1.12										
2002	10	60	50	31	34.1	40.4	36.9	1.93	3.5	0.92	1	0.92										
2001	18	54	40	27	28.7	38.8	33.6	3.16	8.3	0.83	5	0.83										
2000	3	46	51	39	30.7	37.8	31.3	7.23	9.5	0.84	5.5	0.84										
1999	9	86	81	23	27	40.8	39.9	7.01	4.3	2.82	1.3	2.82										
1998	14	85	47	29	34.2	49.8	34.0	5.2	0.5	2.35	0.5	2.35										
1997	4	62	46	17	29.1	39.3	31.6	2.86	4.4	1.28	0.5	1.28										
1996	6	56	40	18	34	37	30.5	3.64	26.1	1.4	18.8	1.4										
1995	10	68	34	27	31.9	43.1	37.5	3.75	12.2	1.45	0.2	1.45										
1994	2	55	35	10	17.9	33.2	25.5	5.62	12	1.87	4.5	1.87										
1993	11	84	41	27	30.1	42.5	36.5	3.44	1.3	1.42	1.1	1.42										
1992	11	62	42	23	28.9	42.5	35.7	1.68	1.5	0.5	1.3	0.5										
1991	10	51	43	20	28.5	41.2	34.9	3.48	8.4	0.85	0.7	0.85										
1990	24	48	48	30	30.2	47.0	42.4	3.54	0.9	1.11	0.8	1.11										
1989	11	61	44	26	32.7	44	36.4	2.29	5	0.4	1	0.4										
1988	1	51	41	27	27.1	35.8	29.5	3.48	13.9	0.8	0.2	0.8										
1987	8	37	46	17	21.2	31.5	22.5	5.41	12.6	1.38	1.2	1.38										
1986	8	61	22	21.4	40.8	40.1	4.91	2.2	2.18	1.4	2.18	1.4										
1985	2	58	40	9	13	34.6	28.8	1	8.4	0.3	4.1	0.3										
1984	8	37	39	39	34.8	37.9	30.9	1.87	11.7	0.4	4.2	0.4										
1983	12	38	43	21	28.4	39.6	34.5	NR	1.9	NR	1	NR										
1982	0	38	37	15	19.6	31.5	28.1	4.46	15.9	0.79	1.8	0.79										
1981	3	49	40	14	20.2	32.3	28.2	0.58	8	0.46	4.1	0.46										
1980	17	40	40	27	28	38.4	31.7	1.72	7	0.81	2	0.81										
1979	8	63	50	20	26.9	40.2	35.5	10.52	6.6	3.45	1.1	3.45										
1978	12	58	31	22	33.9	38	27.7	8.27	44	1.4	10.7	1.4										
1977	-2	44	28	12	16.4	27.7	22.2	2.25	1.3	1.44	5	1.44										
1976	13	76	26	20	25.3	34.7	27.4	2.78	5.4	0.8	1.3	0.8										
1975	20	61	43	26	32.3	40.1	37.9	4.15	2	0.8	1.1	0.8										
1974	6	68	45	23	29.1	42.4	35.2	2.8	12.4	1.29	5.8	1.29										
1973	10	68	45	23	29.1	42.4	35.2	2.8	12.4	1.29	5.8	1.29										
1972	5	62	45	15	38	42.1	35	2.41	2.8	0.44	1.8	0.44										
1971	4	54	34	14	28.8	32.1	29.7	11.4	0.28	6.4	0.4	0.28										
1970	3	51	42	14	18.7	31.5	25.1	0.66	8.4	0.19	2.6	0.19										
1969	48	62	34	24	26.3	31.5	24.8	1.1	1	0.48	1	0.48										
1968	-1	50	36	13	20.1	33.3	28.7	2.06	3.6	1.3	1.5	1.3										
1967	13	58	44	25	31.5	40.2	37.3	1.28	1.4	0.82	0.5	0.82										
1966	8	62	32	19	25.7	38.6	32.2	2.65	11.6	0.72	3.8	0.72										
1965	9	61	41	16	23.7	36.6	29.6	3.09	14.8	0.84	6.3	0.84										
1964	8	55	41	22	29.5	41.8	35.6	4.02	13.3	1.39	11.5	1.39										
1963	1	51	39	20	24	36.2	31.1	3.2	5.8	0.8	1.3	0.8										
1962	11	58	41	25	29.1	39.8	32.8	2.42	0.8	1.17	0.4	1.17										
1961	5	55	42	19	21.9	33.5	27.7	1.88	16.7	0.78	5.5	0.78										
1960	27	54	38	28	28.8	36.9	30.9	2.4	0.8	1.8	1.8	1.8										
1959	24	61	38	22	34.3	38	31.2	2.54	1.5	0.83	0.9	0.83										
1958	12	54	39	22	26.3	37.5	31.9	3.79	9.2	0.86	4.5	0.86										
1957	0	60	43	12	22.2	34.8	28.5	5.7	8.9	0.49	2.9	0.49										
1956	24	48	43	28	26.7	37.3	32	1.54	1.2	0.52	0.5	0.52										
1955	13	50	40	24	25.2	36.7	30.9	0.77	2.6	0.51	0.9	0.51										
1954	7	41	31	11	23.1	38.1	30.8	1.85	12.7	0.42	3.4	0.42										
1953	21	37	49	33	32.8	43.4	37.6	4.9	4.1	1.3	2	1.3										
1952	8	60	45	23	28.9	43.4	36.2	4.88	6.1	0.83	1.8	0.83										
1951	11	44	47	27	26.4	44.5	36.9	3.31	0.9	1.27	0.9	1.27										
1950	26	57	36	16	21.2	40.6	34.6	3.8	4.8	0.8	1.4	0.8										
1949	20	61	44	19	26.1	44.4	36.8	3.38	8.8	1.7	1.5	1.7										
1948	0	45	33	18	19.3	31.2	25.2	5.07	15.3	1.23	3.5	1.23										
1947	12	48	42	29	33	41.1	37.1	3.4	1.4	0.88	4	0.88										
1946	3	62	34	26	27	41.2	34.1	1.86	4.2	0.89	1.8	0.89										
1945	3	51	38	15	19.9	31.5	25.2	2.46	12.3	0.77	6.7	0.77										
1944	12	62	38	23	27.8	40	33.9	3.2	4.8	0.96	3.2	0.96										
1943	26	58	29	24	24.3	37.3	30.7	2.86	5.8	1.23	1.1	1.23										
1942	5	57	43	14	23.9	37.5	30.7	2.88	6.4	1.09	2.8	1.09										
1941	10	47	36	25	23.2	36.6	29.4	3.23	9.2	1.38	5	1.38										
1940	7	54	29	18	18.9	31	25	2.61	2.5	2.06	1.5	2.06										
1939	6	55	38	16	18.9	31	25	2.61	2.5	2.06	1.5	2.06										
1938	6	55	42	17	23.7	38.3	32.1	3.59	6.5	1.35	2.7	1.35										
1937	25	46	30	16	20.7	40.3	34.7	5.07	4.1	0.87	2.7	0.87										
1936	10	57	36	16	20.7	39.3	32.6	5.07	11.7	2.2	2.7	2.2										
1935	-1	58	47	18	21.8	36.1	29.8	3.91	23.9	1.												
1934	7	41	31	11	23.1	38.1	30.8	1.85	12.7	0.42	3.4	0.42										
1933	21	37	49	33	32.8	43.4	37.6	4.9	4.1	1.3	2	1.3										
1932	8	60	45	23	28.9	43.4	36.2															