

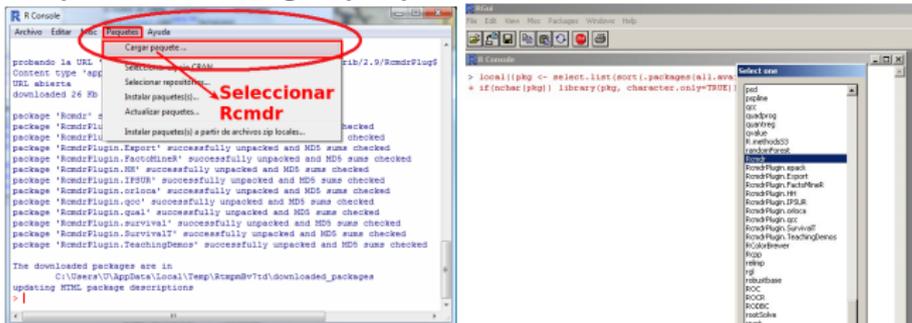
# Práctica 1. Estadística descriptiva

Departamento de Estadística  
Universidad de Oviedo

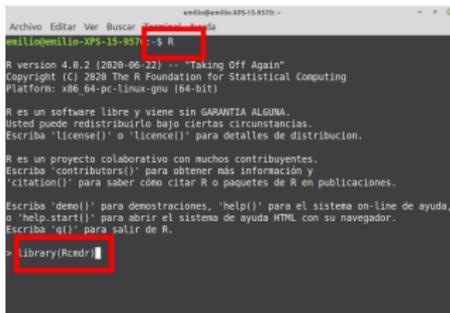
- 1 Cargar la base de datos
- 2 Frecuencias y porcentajes
- 3 Gráficas
- 4 Medidas de centralización y dispersión
- 5 Recodificar y filtrar

# Cargar el programa R y el paquete Rcommander

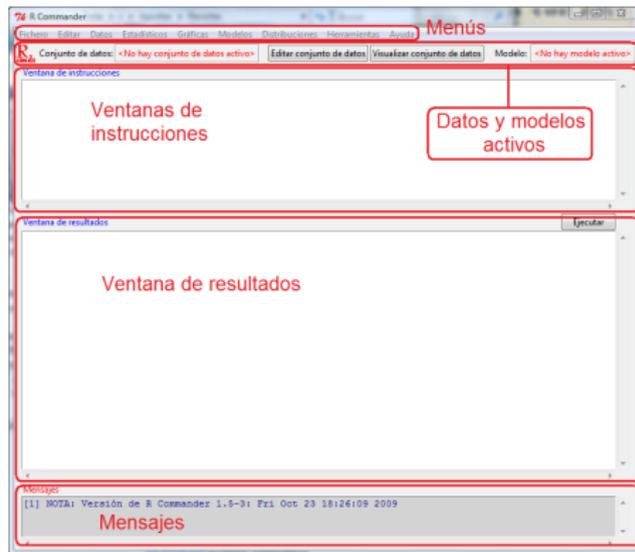
- Iniciamos el programa R.
- Cargamos el paquete RCommander. Dos opciones:
  - 1 Menú *Paquetes* → *Cargar paquete* → Seleccionamos **Rcmdr**.



- 2 Escribimos **library(Rcmdr)** en la consola y pulsamos retorno de carro.



# Cargar Rcommander

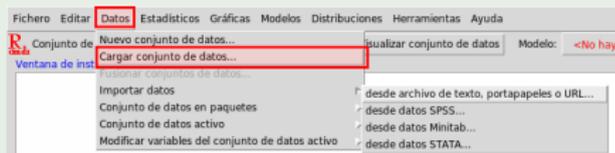


El fichero `acero.rda` se encuentra en el Campus Virtual. Hay que haberlo descargado previamente.

## Cargar la base de datos `acero.rda`

Datos

- ➔ Cargar conjunto de datos
- ➔ Seleccionar **acero.rda**



```
> load("/home/emilio/clases/acero.rda")
```

NOTA: El conjunto de datos `acero` tiene 117 filas y 20 columnas.

- Para visualizar la base de datos:

R Commander interface showing the 'acero' dataset loaded. The 'Visualizar conjunto de datos' button is highlighted with a red box. A red arrow points from this button to a preview window of the data.

	consumo	pr.tbc	pr.cc	pr.ca	pr.galv1	pr.galv2	pr.pint	linea	averia	hora	r
1	135.311	6840	830	0	579	1401	0	1	Si	1	
2	84.082	443	903	58	611	1635	717	1	No	2	
3	131.615	7270	572	36	982	1969	243	1	No	3	
4	90.460	5031	694	122	896	1568	0	1	No	4	
5	120.043	9365	1054	157	403	1480	0	1	No	5	
6	153.678	9281	1003	172	605	1525	473	1	Si	6	
7	99.089	3223	1118	0	643	1424	732	1	No	7	
8	226.375	10490	1077	179	737	1333	93	1	No	8	
9	140.068	7394	1204	167	580	924	247	1	No	9	

Aparece una ventana con los datos disponibles. Moviendo el cursor hacia la izquierda o hacia abajo podemos recorrer toda la base de datos.

Preview window showing a table with 20 variables (columns) and 117 observations (rows).

Tiene 20 variables (columnas)

Dispone de 117 observaciones (filas)

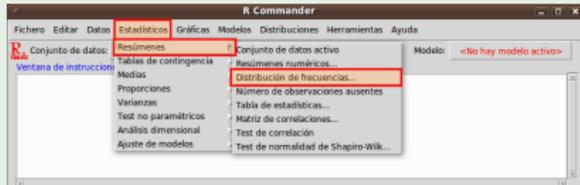
# Variables de la base de datos `acero`

- 1 `consumo` Consumo energético de la empresa (Megavatios/hora).
- 2 `pr.tbc` Producción del tren de bandas calientes (Toneladas de acero).
- 3 `pr.cc` Producción de colada continua (Toneladas de acero).
- 4 `pr.ca` Producción del convertidor de acero (Toneladas de acero).
- 5 `pr.galv1` Producción de galvanizado de tipo I (Tns. de acero).
- 6 `pr.galv2` Producción de galvanizado de tipo II (Tns. de acero).
- 7 `pr.pint` Producción de chapa pintada (Tns. de acero).
- 8 `linea` Línea de producción empleada (A o B).
- 9 `turno` Turno de mañana (M), tarde (T), noche (N).
- 10 `temperatura` Temperatura del sistema: Alta, Media y Baja.
- 11 `pres.aver` Presencia de averías: hubo Averías (A), no hubo averías (NoA).
- 12 `nun.aver` Número de averías detectadas.
- 13 `sistema` Activación de un sistema de detección de sobrecalentamiento: encendido (ON), apagado (OFF).
- 14 ...



# Halle la distribución de frecuencias de la variable `pres.aver`.

- Estadísticos
  - ➔ Resúmenes
    - ➔ Distribución de frecuencias



- Seleccionar la variable `pres.aver`
  - ➔ Aceptamos



```
counts:
pres.aver
  A NoA
28  89
```

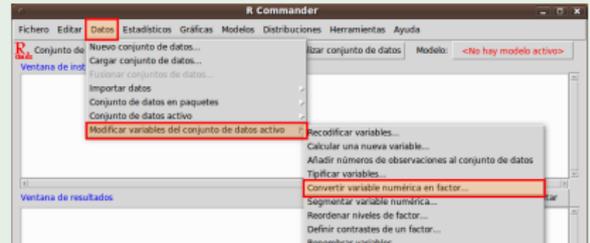
```
percentages:
pres.aver
  A  NoA
23.93 76.07
```

Así, se ha obtenido el número de casos de cada modalidad y el porcentaje que representan dentro de la muestra. La moda es el dato que más se repite; en este caso, la modalidad `NoA`.

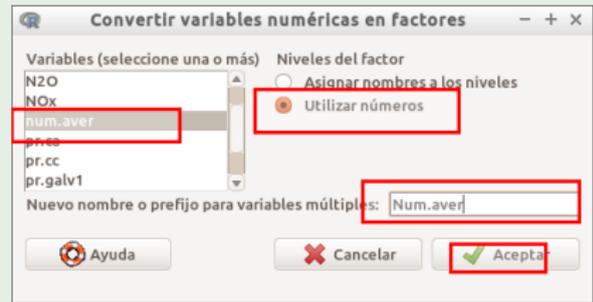


# Convierta la variable *Número de averías* (`num.aver`) en una variable de tipo factor con el nombre `Num.aver`

- 1 Es una variable numérica.
- 2 Datos
  - ↳ Modificar variables del conjunto...
  - ↳ Convertir variable numérica en factor



- 3 Seleccionar la variable `num.aver`
  - ↳ Utilizar números
  - ↳ **Nuevo nombre:**  
**Num.aver**
  - ↳ Aceptar



```
> acero$Num.aver <- as.factor(acero$num.aver)
```

# Obtenga la distribución de frecuencias de la nueva variable

Num.aver.

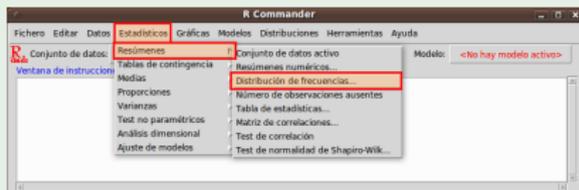
- Estadísticos

  - ➔ Resúmenes

    - ➔ Distribución de frecuencias

- Seleccionar la variable **Num.aver**

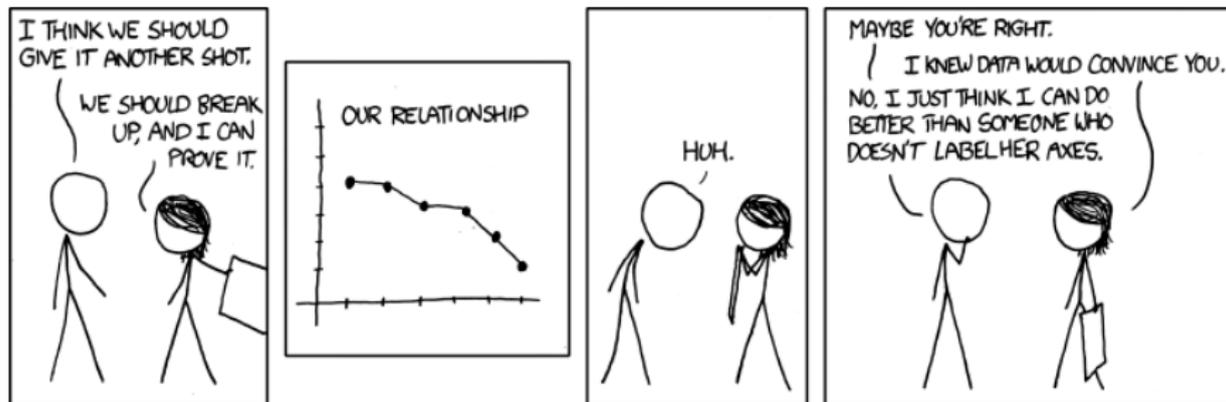
  - ➔ Aceptamos



```
counts:
Num.aver
 0  1  2  3  4
89  2  9  9  8
```

```
percentages:
Num.aver
 0    1    2    3    4
76.07 1.71 7.69 7.69 6.84
```

Un gráfico erróneo es garantía de fracaso.



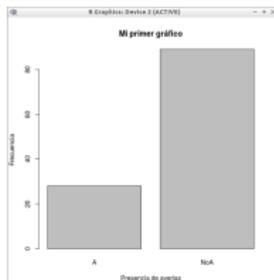
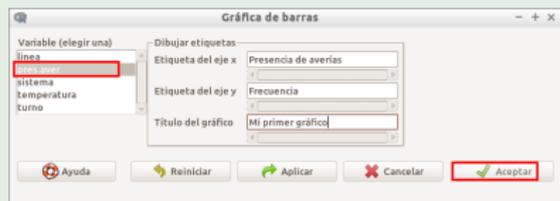
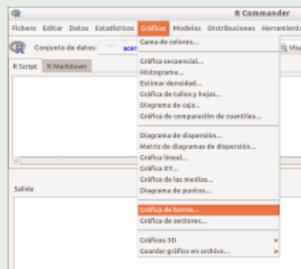
<http://imgs.xkcd.com/comics/convincing.png>

# Represente gráficamente la distribución de la variable `pres.aver` mediante una gráfica de barras.

- Gráficas
  - ➔ Gráfica de barras

Seleccionar la variable **pres.aver**

- y añadimos las etiquetas del gráfico
  - ➔ Aceptamos

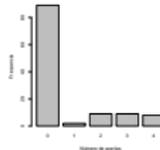
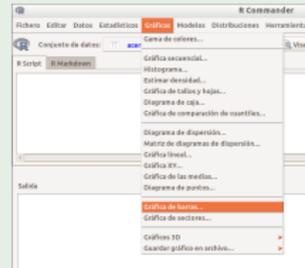


# Obtenga el gráfico de barras de la variable `naverias`

- Es variable numérica, hay que pasarla a factor.
- Trabajaremos con la variable **Num.aver.**

## Gráficas

### ➔ Gráfica de barras



Represente gráficamente la distribución de la variable *averias* mediante un gráfico de sectores.

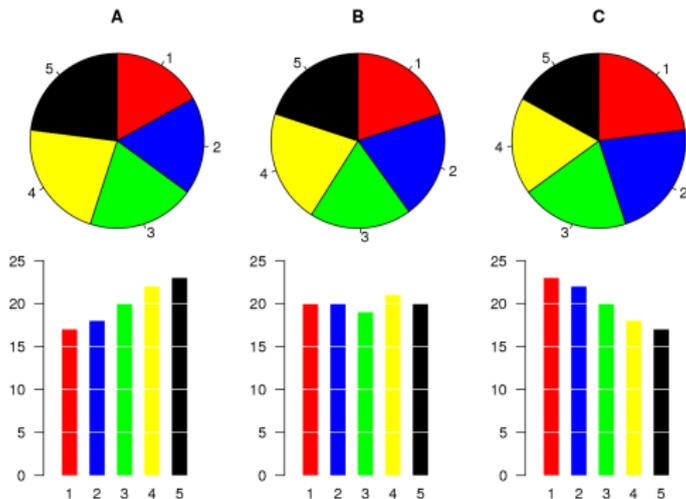
Gráficas

↳ Gráfica de sectores

↳ Seleccionar la variable **pres.aver**

↳ Aceptamos

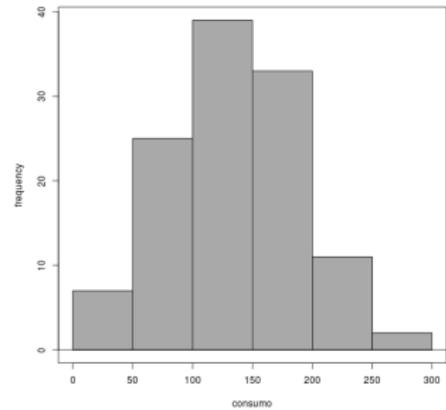
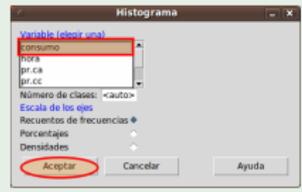
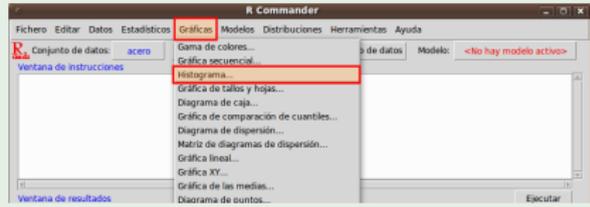
No se recomienda este tipo de gráficos.



# Obtenga el histograma de la variable consumo.

● Gráficas  
    ➡ Histograma...

● Seleccionar la variable **con-**  
**sumo**  
    ➡ Aceptamos



# Obtenga el diagrama de cajas de la variable consumo.

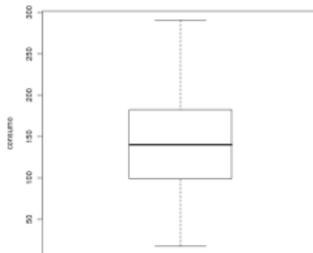
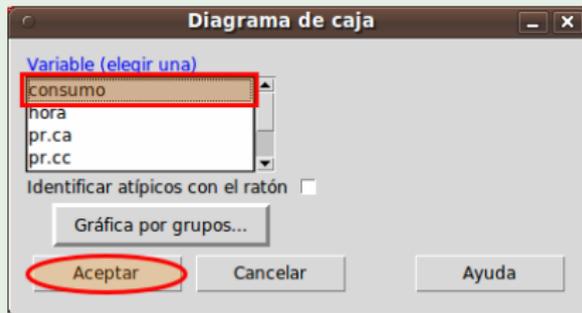
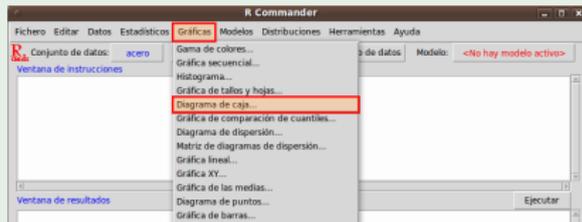
Gráficas

➔ Diagrama de caja...

Seleccionar la variable **con-**

**sumo**

➔ Aceptamos



# Obtenga los diagramas de cajas de consumo según la temperatura.

## Gráficas

↳ Diagrama de caja...

↳ Seleccionar la variable **consumo**

↳ Pulsamos **Gráfica por grupos...**

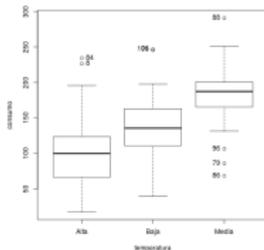


↳ Elegimos el factor **temperatura**

↳ Aceptamos

↳ El botón cambia a *Gráfica según: temperatura*

↳ Aceptamos



# Calcule la media, desviación típica y percentiles de la variable *Número de averías* (`num.aver`).

## Estadísticos

- ➔ Resúmenes
  - ➔ Resúmenes numéricos

## Seleccionar `num.aver`

- ➔ Aceptamos



mean	sd	IQR	0%	25%	50%	75%	100%	n
0.6752137	1.292078	0	0	0	0	0	4	117

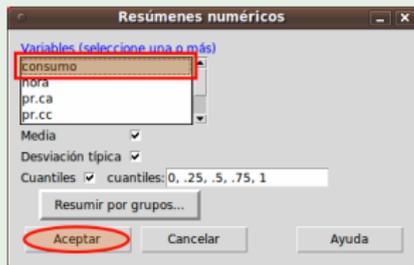
Los resultados nos indican que la media es de 0.67 averías por hora, con una desviación típica de 1.29. El número de averías varía desde 0 hasta 4, y al menos el 75% de la observaciones no presentaron averías. En total disponemos de 117 observaciones.

# Calcule los principales estadísticos descriptivos de la variable consumo.

## Estadísticos

- Resúmenes
  - Resúmenes numéricos

- Seleccionar la variable **consumo**
  - Aceptamos



mean	sd	IQR	0%	25%	50%	75%	100%	n
135.6771	56.90756	83.39	17.5	99.09	135.1	182.48	290.72	117

El consumo medio se sitúa en torno a  $\bar{x}$  ? megavatios-hora, con una desviación típica de  $s$  ? MWh. El consumo mínimo desciende hasta  $x_0$  ? y el máximo asciende hasta  $x_{100}$  ?. El 25% de los casos analizados consumen  $x_{25}$  ? MWh o menos; el 50%, menos de  $x_{50}$  ?; y un  $x_{75}$  % consume más de 182.48.

# Calcule los principales estadísticos descriptivos de la variable consumo para cada nivel de la temperatura.

## Estadísticos

↳ Resúmenes

↳ Resúmenes numéricos

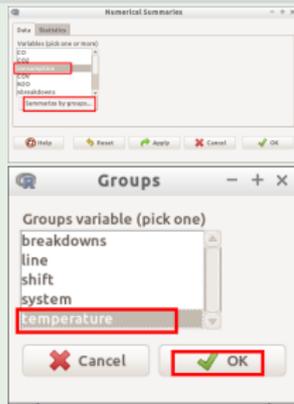
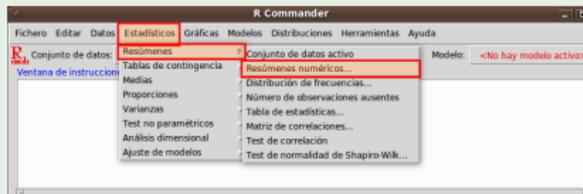
Seleccione la variable consumo

Pinche *Resumir por grupos*

↳ Seleccionar temperatura

↳ OK (en la ventana de Grupos)

OK (en Resúmenes numéricos)



	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Alta	103.5239	51.13951	55.195	17.50	66.4700	99.75	121.6650	234.39	46
Baja	135.5716	46.97305	49.580	39.72	113.1625	135.51	162.7425	246.54	38
Media	180.6182	44.06330	35.190	68.30	165.5600	187.36	200.7500	290.72	33

Estos son las calificaciones del examen de prácticas de laboratorio de Estadística del año pasado. Determine los principales estadísticos.

```
      mean      sd   IQR   0%    25%   50%    75% 100%   n NA
6.728571 2.420019 3.025 0.85 5.7625 7.05 8.7875   10 70 21
```



# Cree una nueva variable GrupoConsumo

$$\text{Grupoconsumo} = \begin{cases} \text{Bajo} & \text{si consumo} \leq 100 \\ \text{Medio} & \text{si } 100 < \text{consumo} \leq 200 \\ \text{Alto} & \text{si consumo} > 200 \end{cases}$$

- Datos

➡ Modificar variables del conjunto de datos activo...

➡ Recodificar variables

- Variable a recodificar **consumo**

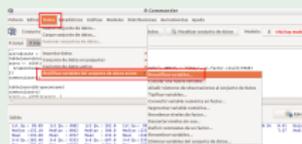
Nueva variable: **Grupoconsumo**

Introducir directrices:

**lo:100= "Bajo"**

**100:200 = "Medio"**

**200:hi = "Alto"**



```
> acero <- within(acero, {  
+   Grupoconsumo <- Recode(consumo, 'lo:100= "Bajo"; 100:200 = "Medio"; 200:hi = "Alto', as.factor  
+ })
```

# Cree una nueva base de datos con los registros que corresponden a la temperatura Alta

- Datos

- ↳ Conjunto de datos activo. . .

- ↳ Filtrar el conjunto de datos activo

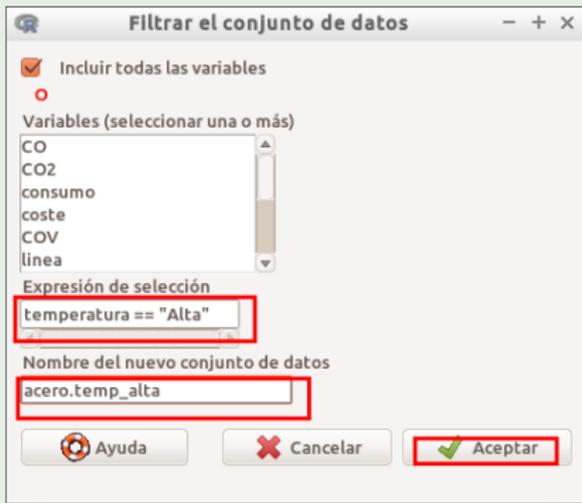
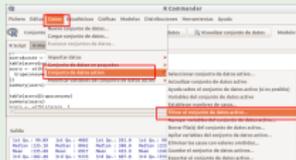
- Expresión de selección:

**temperatura == "Alta"**

Nombre del nuevo conjunto de datos:

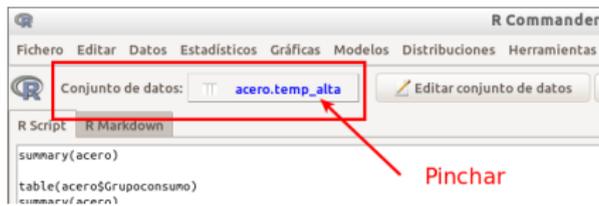
**acero.temp\_alta**

Aceptar



```
> acero.temp_alta <- subset(acero, subset=temperatura == "Alta")
```

NOTA: El conjunto de datos `acero.temp_alta` tiene 46 filas y 23 columnas.



Si queremos activar otras bases:

