# 12 R Programming







## What is R & RStudio?

R is a free, open-source *statistical programming language* and software package for data analysis, visualization, and modeling.

*Interfacing* with other languages such as Python, C, or Fortran is possible, as well as *wrapping* other programs within R scripts.

There is a wide range of options to get support on R, including the extensive R *documentation*, the R *community*, and *commercial support*.

RStudio is a free, open-source *integrated development environment (IDE)* for R.

# Downloading & Installation

Download R

1. Go to https://www.r-project.org
2. Click on the *Download R* link
3. Select a mirror close to your location, or the *0-Cloud* first one to get automatically redirected.
4. Click on the *Download R for Windows* link
5. Click on the *base* link
6. Click on the *Download R 4.x.x for Windows* link
7. Save the *.exe* file in a folder

Install R

1. Double-click the *.exe* file
2. Accept the *default options* in the installation wizard







# Downloading & Installation (cont'd)

Download RStudio

1. Go to https://www.posit.co
2. Click on the *Download RStudio* link
3. Click on the *Download RStudio* link under *RStudio Desktop* option
4. Click on the *Download RStudio Desktop for Windows* link
5. Save the *.exe* file in a folder

Install RStudio

1. Make sure you have the latest version of Java installed
2. Double-click the *.exe* file
3. Accept the *default options* in the installation wizard

# Downloading & Installation (cont'd)

Install the `qcc` package

1. Go to *Tools* tab on the top pane
2. Click on the *Install Packages*; A dialog box appears
3. Type `qcc` on the Packages text box
4. Click on Install

Packages need to be *loaded in the workspace in every session* that uses functions of the package.

Similarly, other packages, such as `SixSigma`, `qualityTools`, `tolerance`, `AcceptanceSampling`, etc., can be installed.



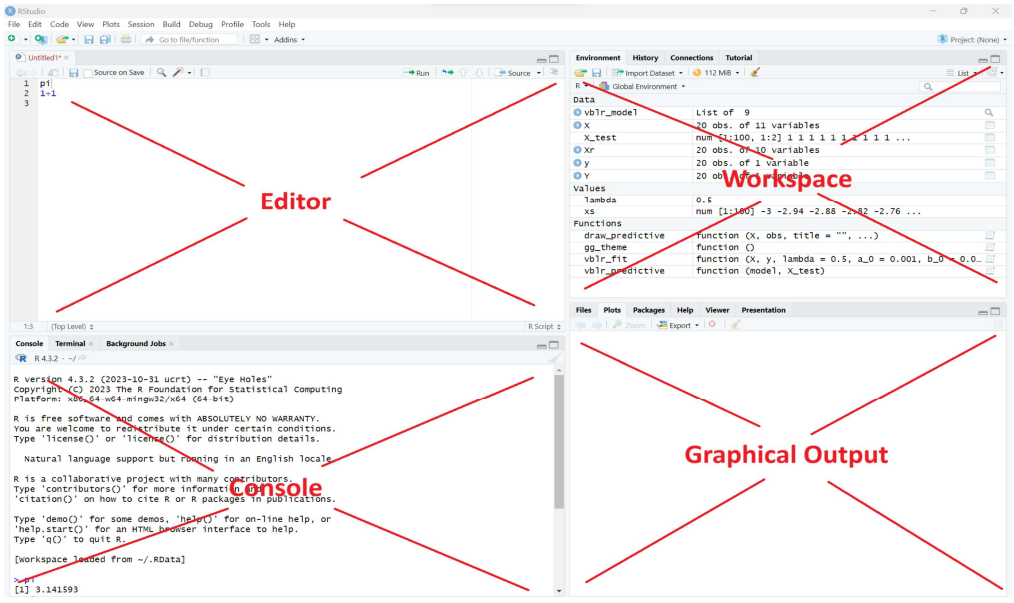




# R Infrastructure

The `R` infrastructure is composed of the following elements:

1. Console
2. Editor
3. Graphical Output
4. Workspace
5. History
6. Working Directory



# Useful R Commands

## General

`;` Separate expressions (in same line)
`<-` Assignment operator

## Math Operators

`+`, `-`, `/`, `*` Arithmatic
`<`, `>`, `<=`, `>=`, `==`, `!=`, `%in%` Comparisons
`&`, `&&`, `|`, `||`, `!` Logic operations

## Integer Operations

`%/%` Integer division
`%%` Module (remainder of a division)







# Useful R Commands (cont'd)

## Math Functions

`sqrt()` Square root
`exp()`, `log()` Exponential and logarithmic
`sin()`, `cos()`, `tan()` Trigonometry
`asin()`, `acos()`, `atan()` Inverse trigonometry
`abs()` Absolute
`round()`, `floor()`, `ceiling()` Rounding
`min()`, `max()` Minimum and maximum
`sum()`, `prod()` Sums and products
`cumsum()`, `cumprod()` Cumulative operations
`factorial()` Factorial

## Vectors

`c()` Create a vector (combine values)
`length()` Vector length
`sort()` Sorting

# Useful R Commands (cont'd)

## Matrices

`matrix()` Create a matrix
`%*%` Matrix multiplication
`t()` Matrix transposition
`solve()` Matrix inversion
`colSums`, `rowSums` Sum by rows or columns
`colMeans`, `rowMeans` Average by rows or columns
`dim`, `nrow`, `ncol` Dimensions

## Files

`read.table()` Import data
`read.csv()` Import data from `.csv` file
`write.csv()` Save `.csv` data file
`save()` Save an `R` data file
`load()` Load an `R` data file into the workspace







# Useful R Commands (cont'd)

## Descriptive Statistics

`mean()` Average
`median()` Median
`quantile()` Quantiles, Percentiles
`var()` Variance
`sd()`, `Standard deviation` Sum by rows or columns

# Pellets Density Example (cont'd)

A certain ceramic process produces pellets whose *density is a critical quality characteristic* according to customer needs.

Current technical specification states that the density of a pellet is considered *acceptable if* it is greater than $10.5$ g/cm$^3$.

A sample of *one pellet* is taken and measured, following a standardized inspection process, after *each hour* of continuous operation.

Pellets density data in g/cm$^3$:

| | | | | | |
|---|---|---|---|---|---|
| 10.6817 | 10.6040 | 10.5709 | 10.7858 | 10.7668 | 10.8101 |
| 10.6905 | 10.6079 | 10.5724 | 10.7736 | 11.0921 | 11.1023 |
| 11.0934 | 10.8530 | 10.6774 | 10.6712 | 10.6935 | 10.5669 |
| 10.8002 | 10.7607 | 10.5470 | 10.5555 | 10.5705 | 10.7723 |





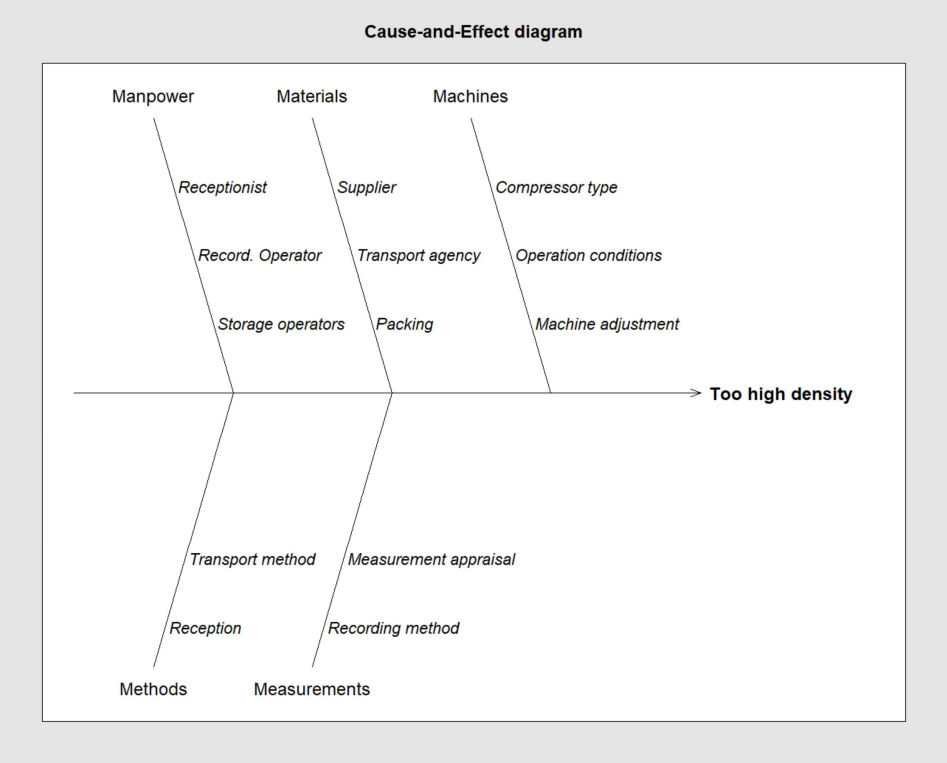


# Cause-and-Effect Diagram

```
cManpower <- c("Recepcionist", "Record. Operator",
               "Storage operators")
cMaterials <- c("Supplier", "Transport agency",
                "Packing")
cMachines <- c("Compressor type",
               "Operation conditions",
               "Machine adjustment")
cMethods <- c("Reception", "Transport method")
cMeasurements <- c("Recording method",
                   "Measurement appraisal")
cGroups <- c("Manpower", "Materials", "Machines",
             "Methods", "Measurements")
cEffect <- "Too high density"
```

```
library(qcc)
cause.and.effect(
  cause = list(Manpower = cManpower,
               Materials = cMaterials,
               Machines = cMachines,
               Methods = cMethods,
               Measurements = cMeasurements),
  effect = cEffect)
```

# Cause-and-Effect Diagram (cont'd)



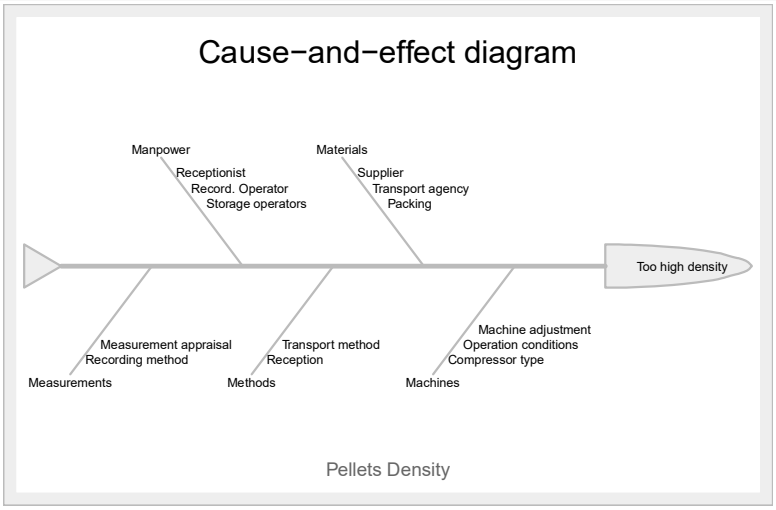







# Cause-and-Effect Diagram (cont'd)

```
library(SixSigma)
ss.ceDiag(
  effect = cEffect,
  causes.gr <- cGroups,
  causes = list(cManpower, cMaterials, cMachines,
                cMethods, cMeasurements),
  main = "Cause-and-effect diagram",
  sub = "Pellets Density")
```



# Control Charts (cont'd)

```
pdensity <- c(10.6817, 10.6040, 10.5709, 10.7858,
              10.7668, 10.8101, 10.6905, 10.6079,
              10.5724, 10.7736, 11.0921, 11.1023,
              11.0934, 10.8530, 10.6774, 10.6712,
              10.6935, 10.5669, 10.8002, 10.7607,
              10.5470, 10.5555, 10.5705, 10.7723)
```

```
myControlChart <- qcc(data = pdensity,
                      type = "xbar.one")
```

```
summary(myControlChart)
```

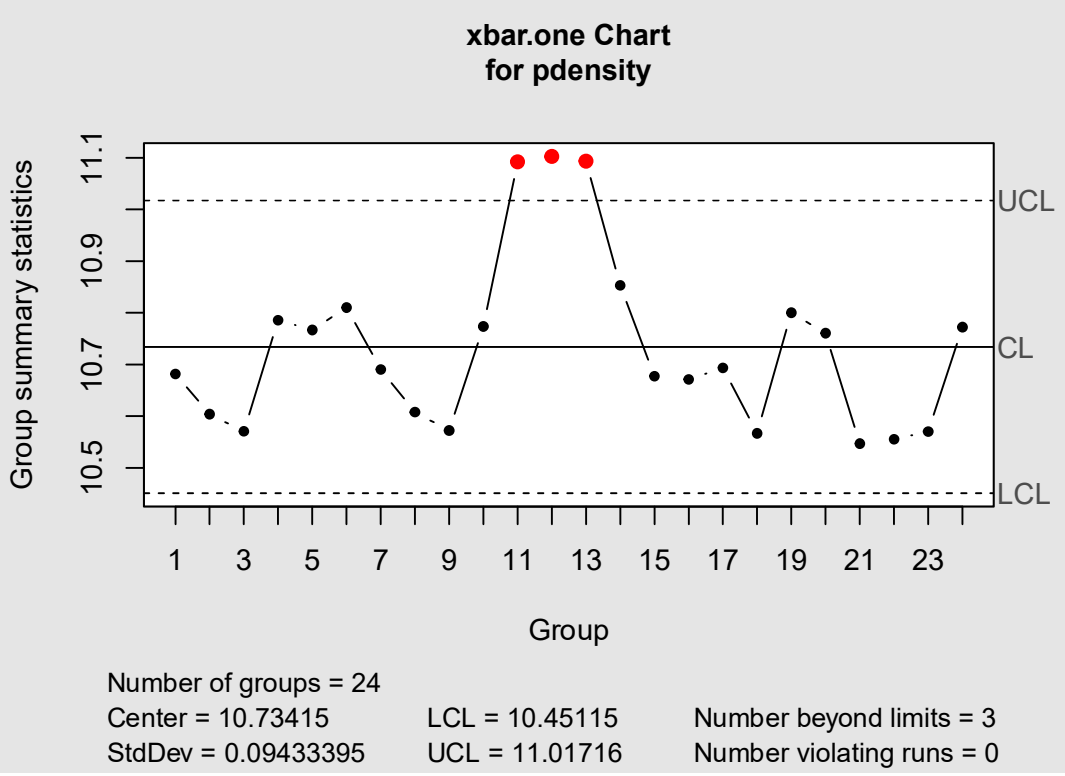
```
##
## Call:
## qcc(data = pdensity, type = "xbar.one")
##
## xbar.one chart for pdensity
##
## Summary of group statistics:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.55   10.60   10.69   10.73   10.79   11.10
##
## Group sample size:  1
## Number of groups:  24
## Center of group statistics:  10.73415
## Standard deviation:  0.09433395
##
## Control limits:
##       LCL      UCL
##  10.45115 11.01716
```





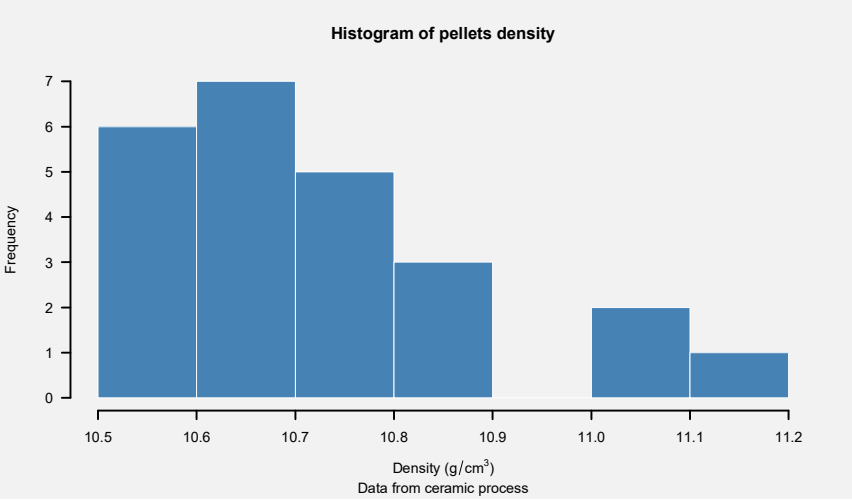


# Control Charts (cont'd)



# Histogram

```
hist(pdensity,
     main = "Histogram of pellets density - Sample #25",
     sub = "Data from ceramic process",
     xlab = expression("Density (g"/"cm"^3*")"),
     col = "steelblue",
     border = "white",
     lwd = 2,
     las = 1,
     bg = "gray")
```










# Check Sheet

## Out of control pellets density check sheet

***Quality Control Department***

***31/01/2015***

Instructions: Mark ticks for the more likely cause of the out-of-control point. Cross every four ticks to make five.

| | Group | Cause | A_supplier | B_supplier | C_supplier |
|---|---|---|---|---|---|
| 1 | Manpower | Recepcionist | // | | |
| 2 | Manpower | Record. Operator | | | / |
| 3 | Manpower | Storage operators | | / | |
| 4 | Machines | Compressor type | || | / | 𝍸 / |
| 5 | Machines | Operation conditions | / | // | |
| 6 | Machines | Machine adjustment | 𝍸 || | / | // |
| 7 | Materials | Supplier | / | 𝍸 𝍸 // | // |
| 8 | Materials | Transport agency | /// | / | //// |
| 9 | Materials | Packing | 𝍸 | // | /// |
| 10 | Methods | Reception | | / | |
| 11 | Methods | Transport method | / | | / |
| 12 | Measurements | Recording method | // | | |
| 13 | Measurements | Measurement appraisal | | / | // |

| Week | Operator | Signature |
|---|---|---|
| 2015-03 | Emilio | |

# Pareto Charts

```
data_checkSheet$A_supplier <- c(2, 0, 0, 2, 1, 7, 1,
                                3, 6, 0, 1, 2, 0)
data_checkSheet$B_supplier <- c(0, 0, 1, 1, 2, 1, 12,
                                1, 2, 1, 0, 0, 1)
data_checkSheet$C_supplier <- c(0, 1, 0, 6, 0, 2, 2,
                                4, 3, 0, 1, 0, 2)
data_checkSheet$Total <- data_checkSheet$A_supplier +
  data_checkSheet$B_supplier +
  data_checkSheet$C_supplier
```

```
data_pareto <- data_checkSheet[order(
  data_checkSheet$Total,
  decreasing = TRUE), ]
par(mar = c(8, 4, 4, 2) + 0.1)
barplot(height = data_pareto$Total,
        names.arg = data_pareto$Cause,
        las = 2,
        main = "Pareto chart for total causes")
```

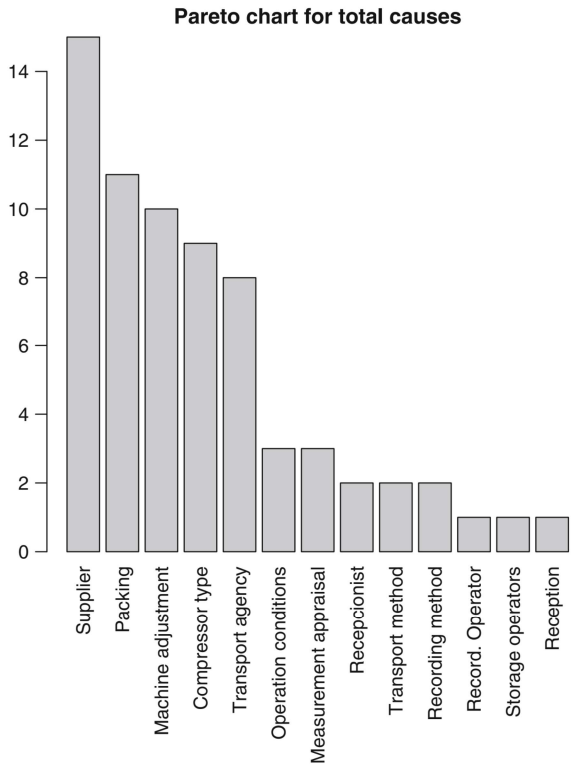
```
library(qcc)
data_pareto2 <- data_pareto$Total
names(data_pareto2) <- data_pareto$Cause
pareto.chart(x = data_pareto2,
             main = "Out-of-control causes")
```



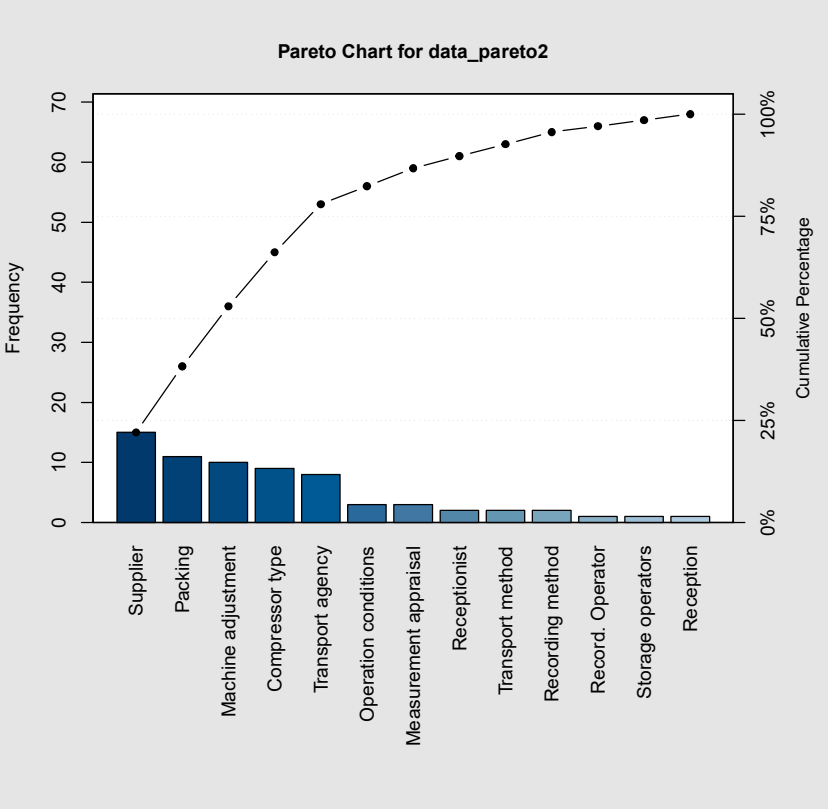




# Pareto Charts (cont'd)



# Pareto Charts (cont'd)



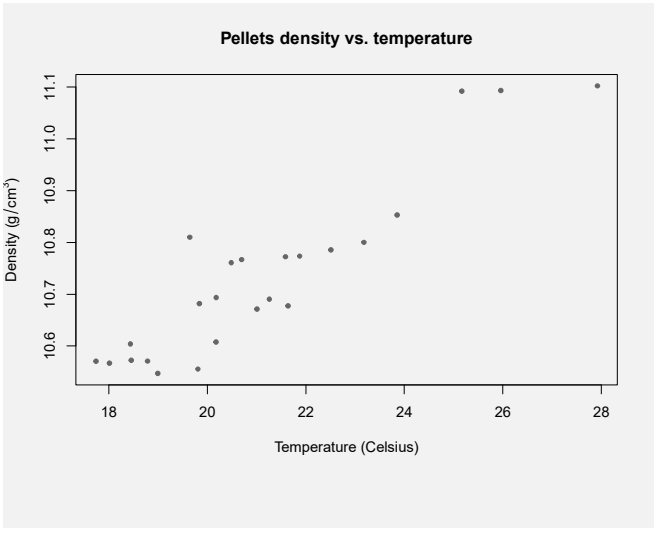







# Scatter Plot

```
set.seed(1234)
ptemp <- - 140 + 15*pdensity + rnorm(24)
plot(pdensity ~ ptemp,
     col = "gray40",
     pch = 20,
     main = "Pellets density vs. temperature",
     xlab = "Temperature (Celsius)",
     ylab = expression("Density ("*g/cm^3*")"))
```



# Box Plot

```
psupplier <- rep(c("A", "B", "C"), each = 8)
boxplot(pdensity ~ psupplier,
        col = "gray70",
        xlab = "Supplier",
        ylab = expression("Density ("*g/cm^3*")"),
        main = "Box plots by supplier")
```