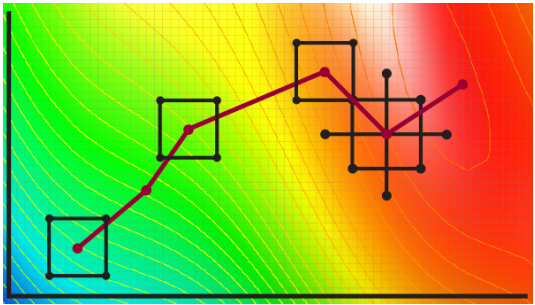
Experimentation for Improvement



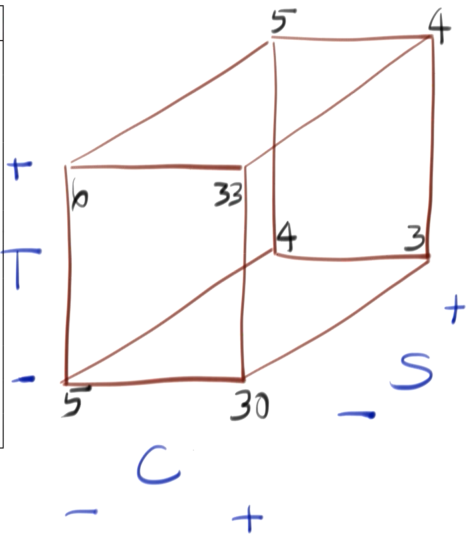
http://learnche.org/

Design and Analysis of Experiments

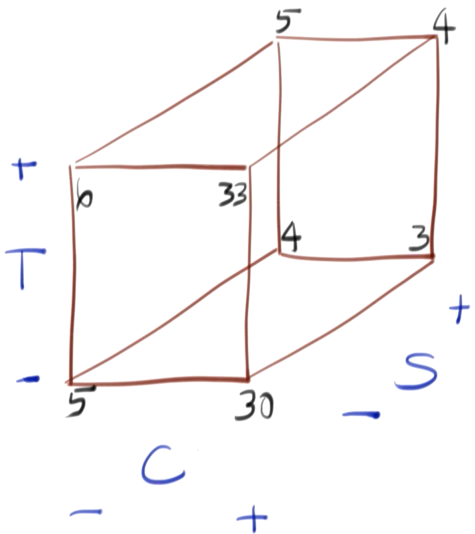
Copyright, sharing, and attribution notice

Waste water treatment example: analysis of the data by hand

Standard order	C	T	S	Outcome [lbs]
1	−	−	−	5
2	+	−	−	30
3	−	+	−	6
4	+	+	−	33
5	−	−	+	4
6	+	−	+	3
7	−	+	+	5
8	+	+	+	4

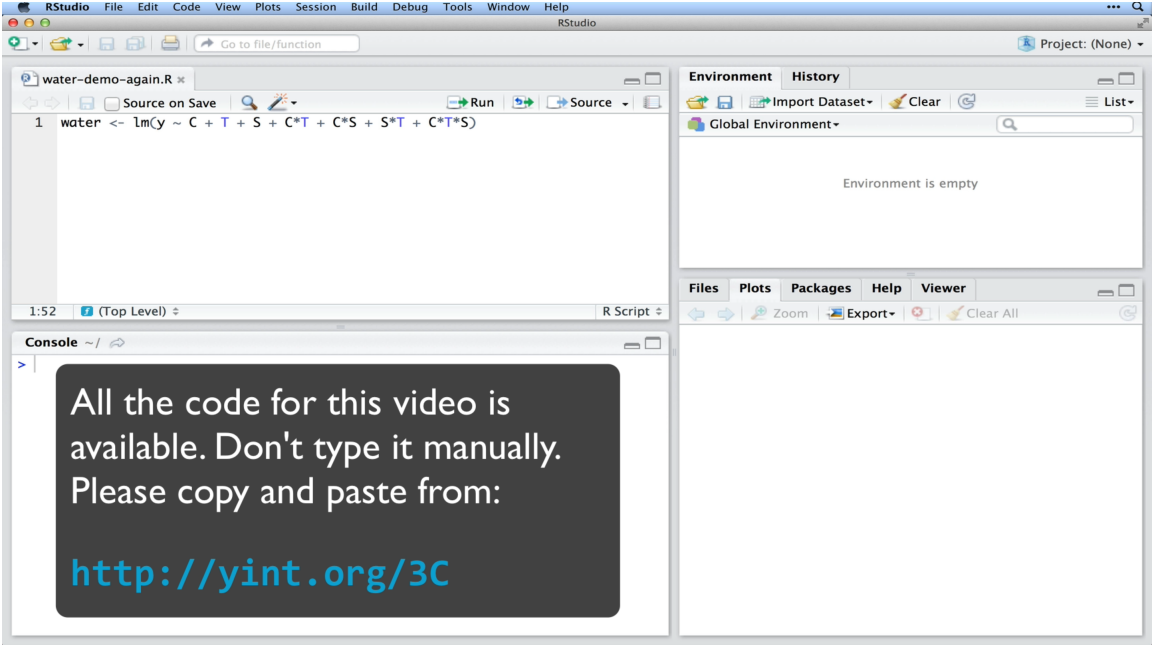


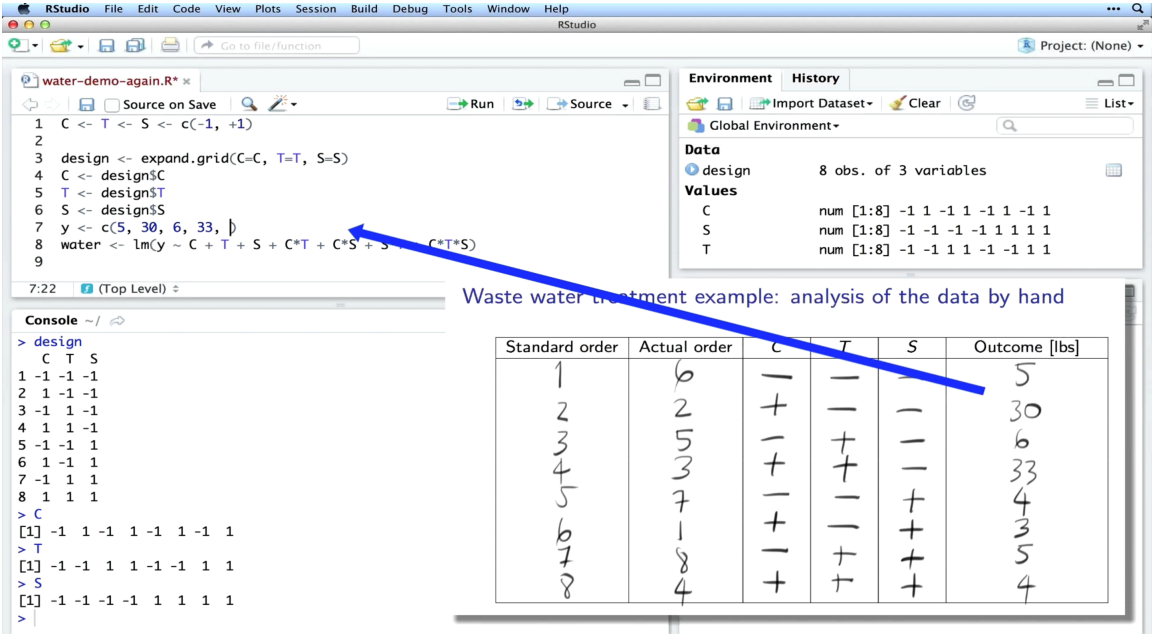
Waste water treatment example: analysis of the data by hand



$$
\begin{aligned}
y = 11.25 \\
&+ 6.25 x_{\mathrm{C}} \\
&+ 0.75 x_{\mathrm{T}} \\
&- 7.25 x_{\mathrm{S}} \\
&- 6.75 x_{\mathrm{C}} x_{\mathrm{S}} \\
&+ 0.25 x_{\mathrm{C}} x_{\mathrm{T}} \\
&- 0.25 x_{\mathrm{T}} x_{\mathrm{S}} \\
&- 0.25 x_{\mathrm{C}} x_{\mathrm{T}} x_{\mathrm{S}}
\end{aligned}
$$

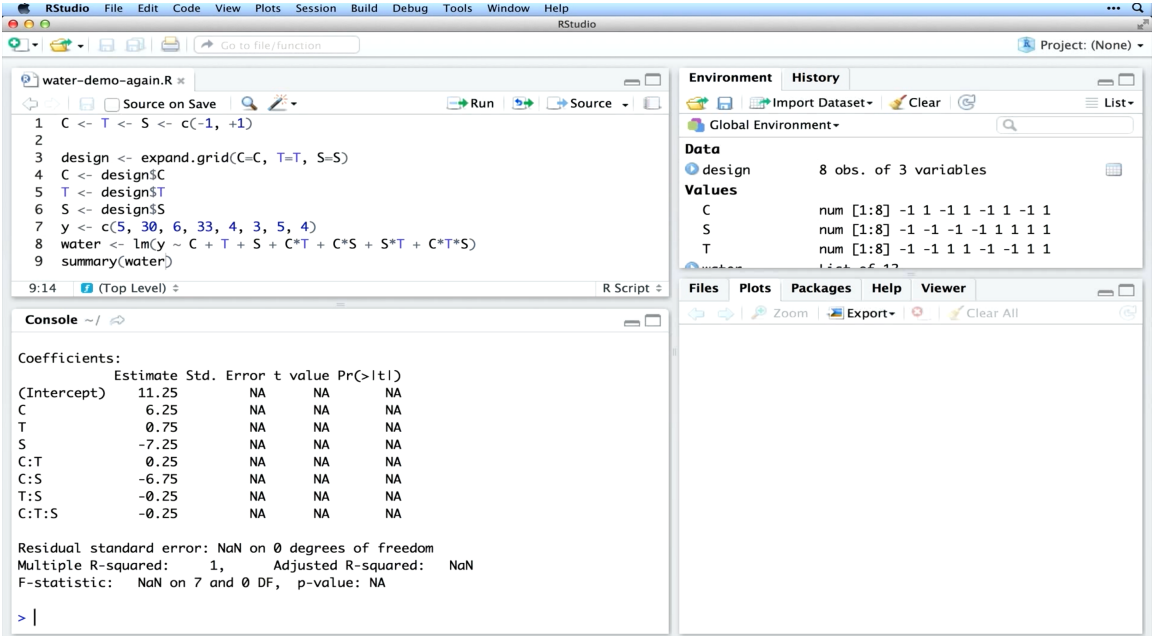
8 parameters


RStudio File Edit Code View Plots Session Build Debug Tools Window Help
RStudio
Go to file/function
Project: (None)
water-demo-again.R
Source on Save
Run
Source
1 water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
1:52
(Top Level)
R Script
Console ~/
>
Environment History
Import Dataset
Clear
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Zoom
Export
Clear All
All the code for this video is available. Don't type it manually. Please copy and paste from:
http://yint.org/3C



Waste water treatment example: analysis of the data by hand

Standard order	Actual order	C	T	S	Outcome [lbs]
1	6	−	−	−	5
2	2	+	−	−	30
3	5	−	+	−	6
4	3	+	+	−	33
5	7	−	−	+	4
6	1	+	−	+	3
7	8	−	+	+	5
8	4	+	+	+	4


RStudio File Edit Code View Plots Session Build Debug Tools Window Help
RStudio
Go to file/function
Project: (None)
water-demo-again.R
Source on Save
Run
Source
1 C <- T <- S <- c(-1, +1)
2
3 design <- expand.grid(C=C, T=T, S=S)
4 C <- design$C
5 T <- design$T
6 S <- design$S
7 y <- c(5, 30, 6, 33, 4, 3, 5, 4)
8 water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
9 summary(water)
9:14
(Top Level)
R Script
Console ~/
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.25 NA NA NA
C 6.25 NA NA NA
T 0.75 NA NA NA
S -7.25 NA NA NA
C:T 0.25 NA NA NA
C:S -6.75 NA NA NA
T:S -0.25 NA NA NA
C:T:S -0.25 NA NA NA
Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: NaN
F-statistic: NaN on 7 and 0 DF, p-value: NA
>
Environment
History
Import Dataset
Clear
List
Global Environment
Data
design 8 obs. of 3 variables
Values
C num [1:8] -1 1 -1 1 -1 1 -1 1
S num [1:8] -1 -1 -1 -1 1 1 1 1
T num [1:8] -1 -1 1 1 -1 -1 1 1
Files
Plots
Packages
Help
Viewer
Zoom
Export
Clear All

RStudio File Edit Code View Plots Session Build Debug Tools Window Help

RStudio

Go to file/function

Project: (None)

water-demo-again.R

Source on Save | Run | Source

```
1  C <- T <- S <- c(-1, +1)
2
3  design <- expand.grid(C=C, T=T, S=S)
4  C <- design$C
5  T <- design$T
6  S <- design$S
7  y <- c(5, 30, 6, 33, 4, 3, 5, 4)
8  water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
9  summary(water)
```

9:14 (Top Level)

Console ~/

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.25         NA      NA       NA
C               6.25         NA      NA       NA
T               0.75         NA      NA       NA
S              -7.25         NA      NA       NA
C:T             0.25         NA      NA       NA
C:S            -6.75         NA      NA       NA
T:S            -0.25         NA      NA       NA
C:T:S          -0.25         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared:      1,    Adjusted R-squared:   NaN
F-statistic:    NaN on 7 and 0 DF,  p-value: NA
```

>

Environment History

Import Dataset | Clear | List

Global Environment

Data

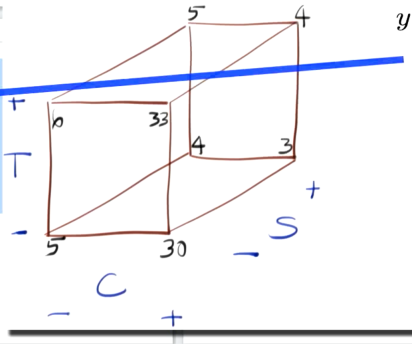
design 8 obs. of 3 variables

Values

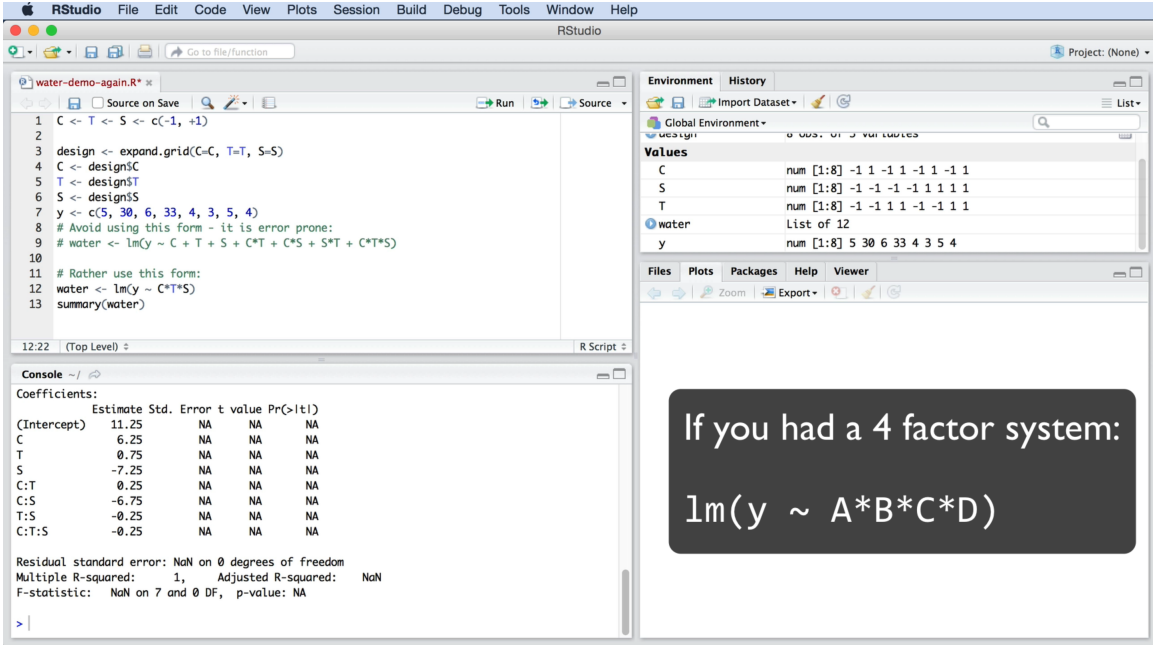
C num [1:8] -1 1 -1 1 -1 1 -1 1

S num [1:8] -1 -1 -1 -1 1 1 1 1

Waste water treatment example: analysis of the data by hand



$$y = 11.25 + 6.25x_C + 0.75x_T - 7.25x_S - 6.75x_Cx_S + 0.25x_Cx_T - 0.25x_Tx_S - 0.25x_Cx_Tx_S$$

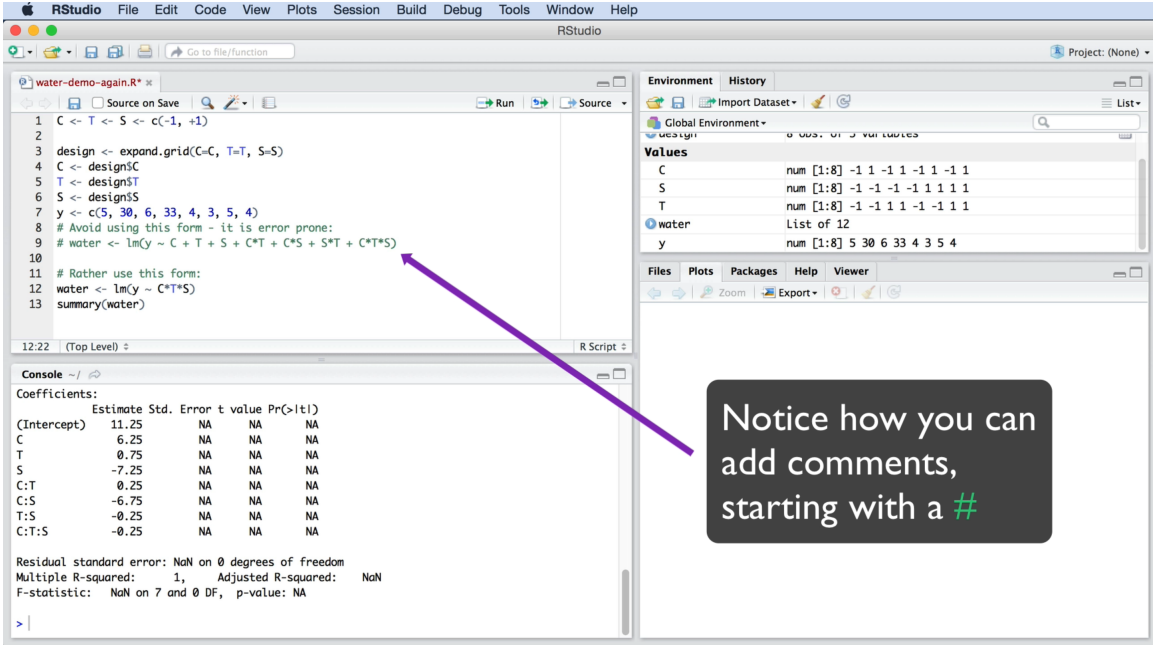

RStudio File Edit Code View Plots Session Build Debug Tools Window Help
RStudio
Go to file/function
Project: (None)
water-demo-again.R*
Source on Save
Run
Source
1 C <- T <- S <- c(-1, +1)
2
3 design <- expand.grid(C=C, T=T, S=S)
4 C <- design$C
5 T <- design$T
6 S <- design$S
7 y <- c(5, 30, 6, 33, 4, 3, 5, 4)
8 # Avoid using this form - it is error prone:
9 # water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
12:22 (Top Level)
R Script
Console ~/
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.25 NA NA NA
C 6.25 NA NA NA
T 0.75 NA NA NA
S -7.25 NA NA NA
C:T 0.25 NA NA NA
C:S -6.75 NA NA NA
T:S -0.25 NA NA NA
C:T:S -0.25 NA NA NA
Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: NaN
F-statistic: NaN on 7 and 0 DF, p-value: NA
>
Environment History
Import Dataset
List
Global Environment
Values
C num [1:8] -1 1 -1 1 -1 1 -1 1
S num [1:8] -1 -1 -1 -1 1 1 1 1
T num [1:8] -1 -1 1 1 -1 -1 1 1
water List of 12
y num [1:8] 5 30 6 33 4 3 5 4
Files Plots Packages Help Viewer
Zoom
Export
If you had a 4 factor system:
lm(y ~ A*B*C*D)

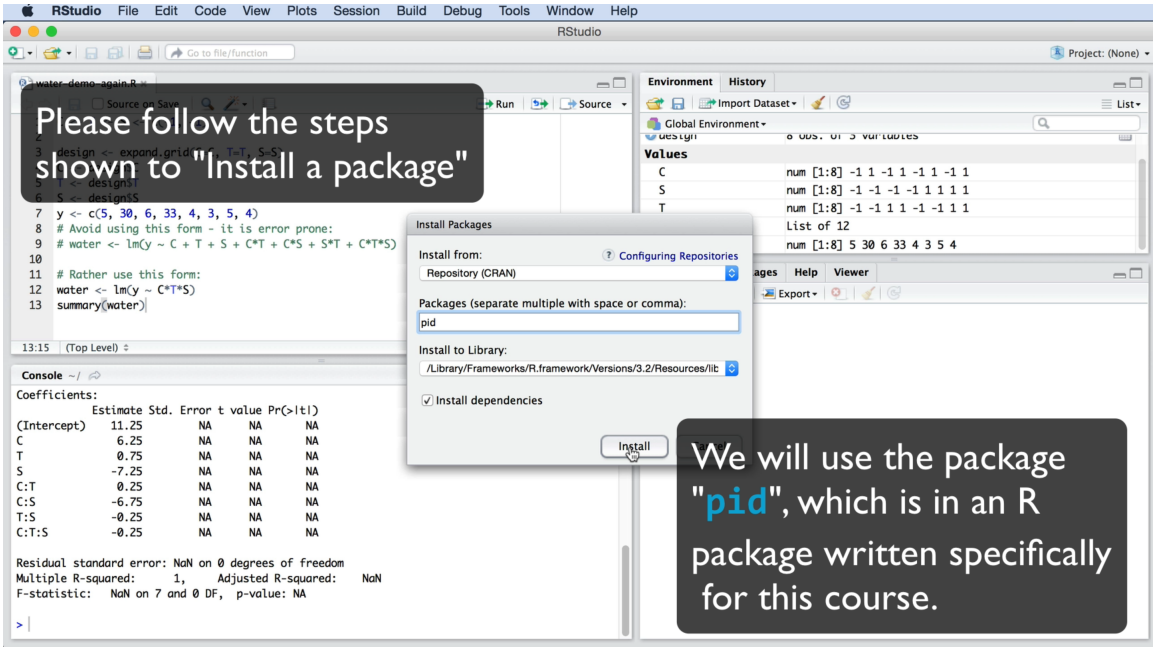
There are many good reasons for writing code for your data analysis

1. It creates a record of your work, detailing the exact steps

2. It is traceable: showing the data source, and how you subsequently analyzed it
   - menu driven software is not traceable in this way

3. You can save it, and share it with colleagues
   - they can reproduce your steps exactly


RStudio File Edit Code View Plots Session Build Debug Tools Window Help
RStudio
Go to file/function
Project: (None)
water-demo-again.R*
Source on Save
Run
Source
1 C <- T <- S <- c(-1, +1)
2
3 design <- expand.grid(C=C, T=T, S=S)
4 C <- design$C
5 T <- design$T
6 S <- design$S
7 y <- c(5, 30, 6, 33, 4, 3, 5, 4)
8 # Avoid using this form - it is error prone:
9 # water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
12:22 (Top Level)
R Script
Console ~/
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.25 NA NA NA
C 6.25 NA NA NA
T 0.75 NA NA NA
S -7.25 NA NA NA
C:T 0.25 NA NA NA
C:S -6.75 NA NA NA
T:S -0.25 NA NA NA
C:T:S -0.25 NA NA NA
Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: NaN
F-statistic: NaN on 7 and 0 DF, p-value: NA
>
Environment History
Import Dataset
List
Global Environment
Values
C num [1:8] -1 1 -1 1 -1 1 -1 1
S num [1:8] -1 -1 -1 -1 1 1 1 1
T num [1:8] -1 -1 1 1 -1 -1 1 1
water List of 12
y num [1:8] 5 30 6 33 4 3 5 4
Files Plots Packages Help Viewer
Zoom
Export
Notice how you can add comments, starting with a #


RStudio File Edit Code View Plots Session Build Debug Tools Window Help
RStudio
Go to file/function
Project: (None)
water-demo-again.R
Source on Save
Run
Source
Please follow the steps shown to "Install a package"
design <- expand.grid(C = T=T, S=S)
C <- design$C
T <- design$T
S <- design$S
y <- c(5, 30, 6, 33, 4, 3, 5, 4)
# Avoid using this form - it is error prone:
# water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
# Rather use this form:
water <- lm(y ~ C*T*S)
summary(water)
13:15 (Top Level)
Console ~/
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.25 NA NA NA
C 6.25 NA NA NA
T 0.75 NA NA NA
S -7.25 NA NA NA
C:T 0.25 NA NA NA
C:S -6.75 NA NA NA
T:S -0.25 NA NA NA
C:T:S -0.25 NA NA NA
Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: NaN
F-statistic: NaN on 7 and 0 DF, p-value: NA
Environment History
Import Dataset
List
Global Environment
Values
C num [1:8] -1 1 -1 1 -1 1 -1 1
S num [1:8] -1 -1 -1 -1 1 1 1 1
T num [1:8] -1 -1 1 1 -1 -1 1 1
List of 12
num [1:8] 5 30 6 33 4 3 5 4
Help Viewer
Export
Install Packages
Install from:
Configuring Repositories
Repository (CRAN)
Packages (separate multiple with space or comma):
pid
Install to Library:
/Library/Frameworks/R.framework/Versions/3.2/Resources/lib
Install dependencies
Install
We will use the package "pid", which is in an R package written specifically for this course.

RStudio File Edit Code View Plots Session Build Debug Tools Window Help

RStudio

Go to file/function

Project: (None)

water-demo-again.R

Source on Save | Run | Source

```
2
3  design <- expand.grid(C=C, T=T, S=S)
4  C <- design$C
5  T <- design$T
6  S <- design$S
7  y <- c(5, 30, 6, 33, 4, 3, 5, 4)
8  # Avoid using this form - it is error prone:
9  # water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
14
15 library(pid)
16 paretoPlot(water)
```

16:18 (Top Level) R Script

Console ~/

```
The following objects are masked from 'package:stats':

    aov, lm

The following object is masked from 'package:graphics':

    plot.design

The following object is masked from 'package:base':

    lengths


> paretoPlot(water)
>
```

Environment History

Import Dataset List

Global Environment

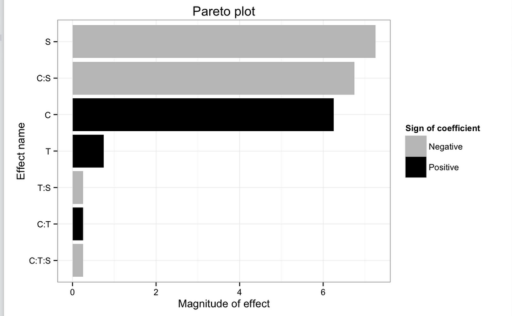
design 8 obs. of 3 variables

Values

C	num [1:8] -1 1 -1 1 -1 1 -1 1
S	num [1:8] -1 -1 -1 -1 1 1 1 1
T	num [1:8] -1 -1 1 1 -1 -1 1 1
water	List of 12
y	num [1:8] 5 30 6 33 4 3 5 4

Files Plots Packages Help Viewer

Zoom Export Publish



RStudio File Edit Code View Plots Session Build Debug Tools Window Help

water-demo-again.R

Source on Save

```
2
3  design <- expand.grid(C=C, T=T,
4  C <- design$C
5  T <- design$T
6  S <- design$S
7  y <- c(5, 30, 6, 33, 4, 3, 5, 4
8  # Avoid using this form - it is
9  # water <- lm(y ~ C + T + S + C
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
14
15 library(pid)
16 paretoPlot(water)
```

13:1 (Top Level)

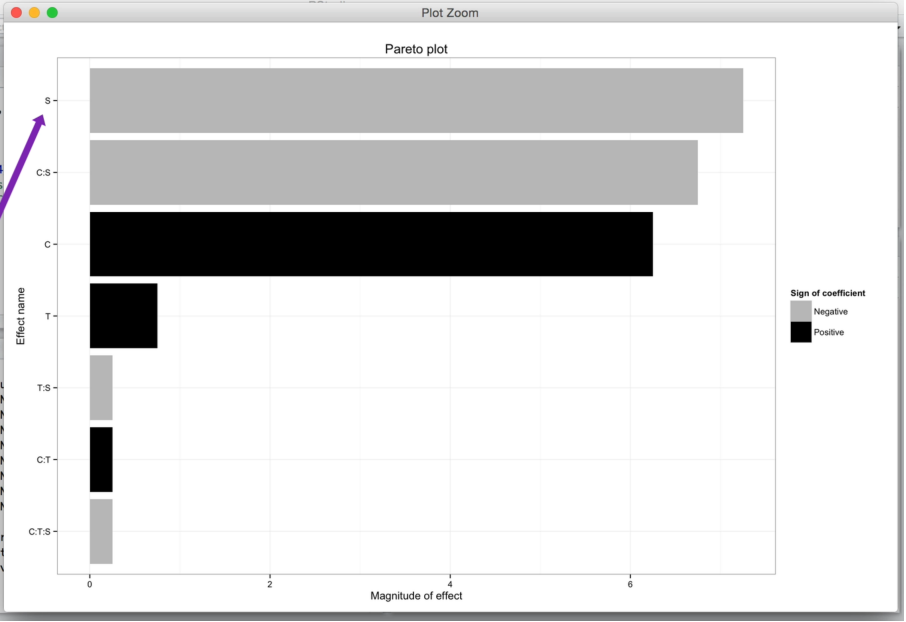
Console ~/

```
Coefficients:
            Estimate Std. Error t valu
(Intercept)    11.25         NA      N
C               6.25         NA      N
T               0.75         NA      N
S              -7.25         NA      N
C:T             0.25         NA      N
C:S            -6.75         NA      N
T:S            -0.25         NA      N
C:T:S          -0.25         NA      N

Residual standard error: NaN on 0 degr
Multiple R-squared:      1,     Adjust
F-statistic:   NaN on 7 and 0 DF,  p-v
```

>

Plot Zoom



RStudio File Edit Code View Plots Session Build Debug Tools Window Help

Plot Zoom

Highest influence

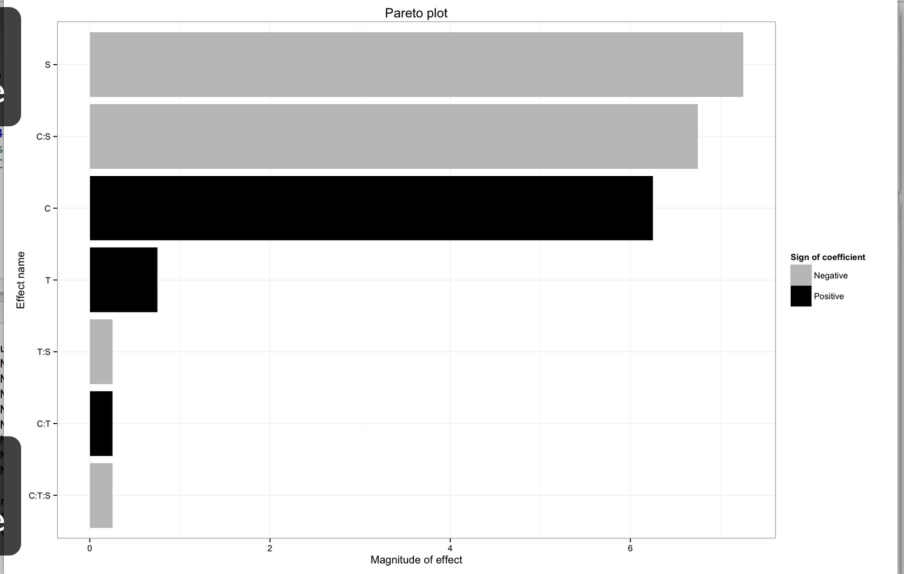
Lowest influence

```
2
3  design <- expand.grid(C=C, T=T,
4  C <- desi
5  T <- desi
6  S <- design$S
7  y <- c(5, 30, 6, 33, 4, 3, 5, 4
8  # Avoid using this form - it is
9  # water <- lm(y ~ C + T + S + C
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
14
15 library(pid)
16 paretoPlot(water)
13:1  (Top Level)
```

Console ~/

```
Coefficients:
             Estimate Std. Error t valu
(Intercept)    11.25         NA
C               6.25         NA
T               0.75         NA
S              -7.25         NA
C:T             0.25         NA
C:S             6.75         NA
T:S             0.25         NA
C:T:S           7.25         NA

Residual standard error: NaN on 0 degr
Multiple R-squared:
F-statistic:
```



RStudio File Edit Code View Plots Session Build Debug Tools Window Help

water-demo-again.R

```
2
3  design <- expand.grid(C=C, T=T,
4  C <- design$C
5  T <- design$T
6  S <- design$S
7  y <- c(5, 30, 6, 33, 4, 3, 5, 4
8  # Avoid using this form - it is
9  # water <- lm(y ~ C + T + S + C
10
11 # Rather use this form:
12 water <- lm(y ~ C*T*S)
13 summary(water)
14
15 library(pid)
16 paretoPlot(water)
```

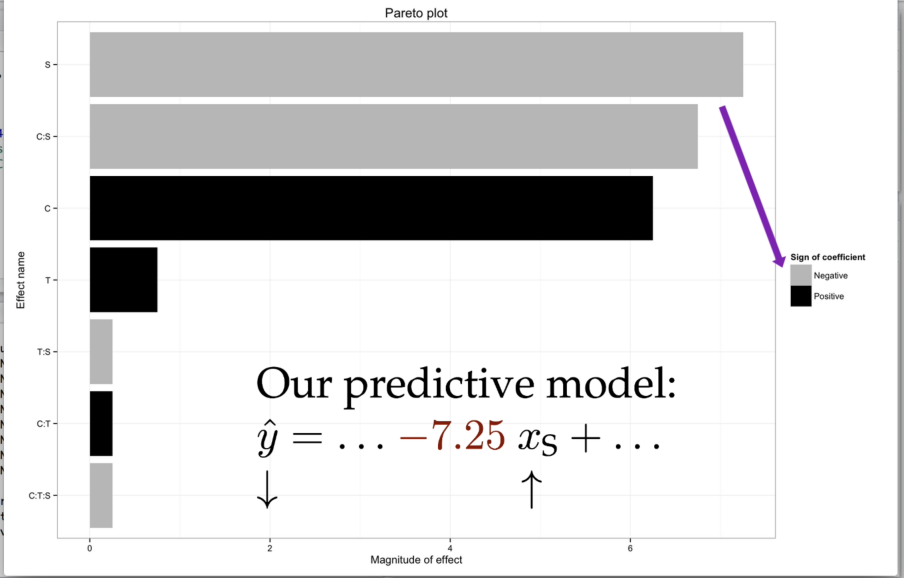
13:1 (Top Level)

Console

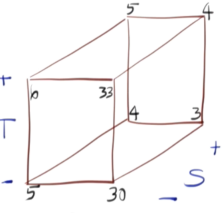
```
Coefficients:
            Estimate Std. Error t valu
(Intercept)    11.25         NA
C               6.25         NA
T               0.75         NA
S              -7.25         NA
C:T             0.25         NA
C:S            -6.75         NA
T:S            -0.25         NA
C:T:S          -0.25         NA

Residual standard error: NaN on 0 degr
Multiple R-squared:      1,     Adjust
F-statistic:   NaN on 7 and 0 DF,  p-v
```

Plot Zoom



Standard order	C	T	S	Outcome [lbs]
1	−	−	−	5
2	+	−	−	30
3	−	+	−	6
4	+	+	−	33
5	−	−	+	4
6	+	−	+	3
7	−	+	+	5
8	+	+	+	4



Our predictive model:

$$\hat{y} = b_0 + b_C\, x_C + b_T\, x_T + b_S\, x_S + b_{CT}\, x_C x_T + b_{CS}\, x_C x_S + b_{TS}\, x_T x_S + b_{CTS}\, x_C x_T x_S$$

$$\begin{aligned}
\text{Experiment } 1: \quad & y_1 = b_0 + b_C(-1) + b_T(-1) + b_S(-1) + \ldots \\
2: \quad & y_2 = b_0 + b_C(+1) + b_T(-1) + b_S(-1) + \ldots \\
3: \quad & y_3 = b_0 + b_C(-1) + b_T(+1) + b_S(-1) + \ldots \\
4: \quad & y_4 = b_0 + b_C(+1) + b_T(+1) + b_S(-1) + \ldots \\
5: \quad & y_5 = b_0 + b_C(-1) + b_T(-1) + b_S(+1) + \ldots \\
6: \quad & y_6 = b_0 + b_C(+1) + b_T(-1) + b_S(+1) + \ldots \\
7: \quad & y_7 = b_0 + b_C(-1) + b_T(+1) + b_S(+1) + \ldots \\
8: \quad & y_8 = b_0 + b_C(+1) + b_T(+1) + b_S(+1) + \ldots
\end{aligned}$$

Our predictive model:

$$\hat{y} = b_0 + b_C\, x_C + b_T\, x_T + b_S\, x_S + b_{CT}\, x_C x_T + b_{CS}\, x_C x_S + b_{TS}\, x_T x_S + b_{CTS}\, x_C x_T x_S$$

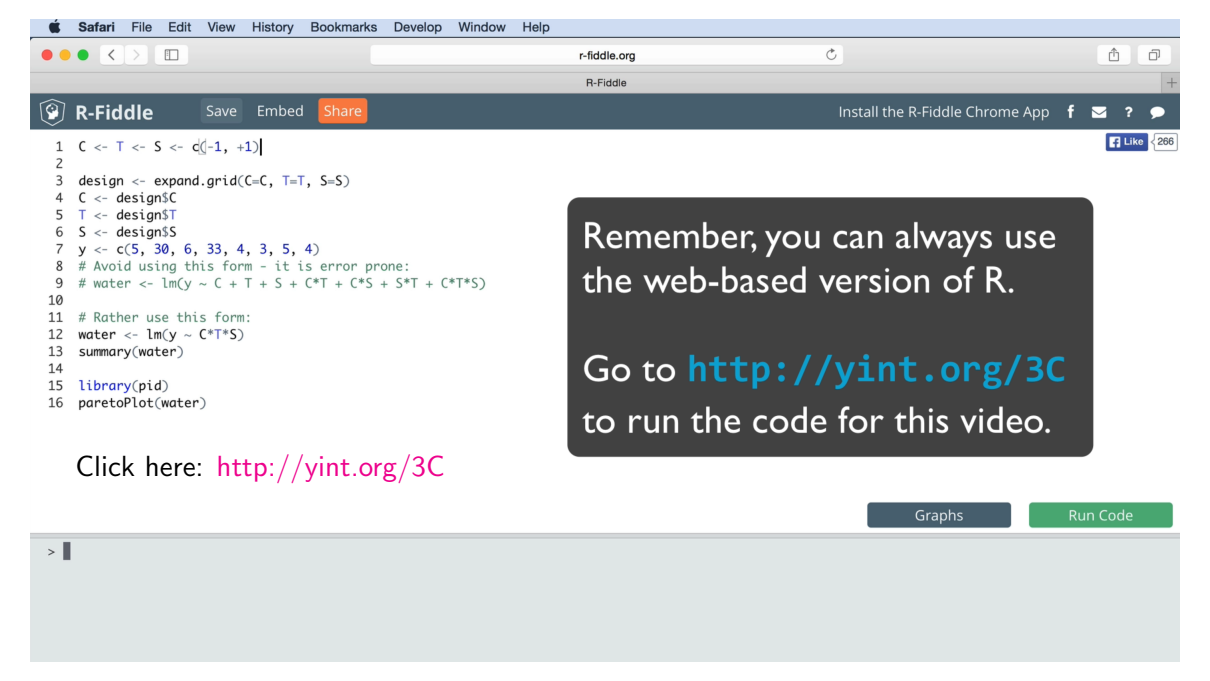
1:	$y_1 = b_0$	+	$b_C(-1)$	+	$b_T(-1)$	+	$b_S(-1)$	+	$b_{CT}(+1)$	+	$b_{CS}(+1)$	+	$b_{TS}(+1)$	+	$b_{CTS}(-1)$
2:	$y_2 = b_0$	+	$b_C(+1)$	+	$b_T(-1)$	+	$b_S(-1)$	+	$b_{CT}(-1)$	+	$b_{CS}(-1)$	+	$b_{TS}(+1)$	+	$b_{CTS}(+1)$
3:	$y_3 = b_0$	+	$b_C(-1)$	+	$b_T(+1)$	+	$b_S(-1)$	+	$b_{CT}(-1)$	+	$b_{CS}(+1)$	+	$b_{TS}(-1)$	+	$b_{CTS}(+1)$
4:	$y_4 = b_0$	+	$b_C(+1)$	+	$b_T(+1)$	+	$b_S(-1)$	+	$b_{CT}(+1)$	+	$b_{CS}(-1)$	+	$b_{TS}(-1)$	+	$b_{CTS}(-1)$
5:	$y_5 = b_0$	+	$b_C(-1)$	+	$b_T(-1)$	+	$b_S(+1)$	+	$b_{CT}(+1)$	+	$b_{CS}(-1)$	+	$b_{TS}(-1)$	+	$b_{CTS}(+1)$
6:	$y_6 = b_0$	+	$b_C(+1)$	+	$b_T(-1)$	+	$b_S(+1)$	+	$b_{CT}(-1)$	+	$b_{CS}(+1)$	+	$b_{TS}(-1)$	+	$b_{CTS}(-1)$
7:	$y_7 = b_0$	+	$b_C(-1)$	+	$b_T(+1)$	+	$b_S(+1)$	+	$b_{CT}(-1)$	+	$b_{CS}(-1)$	+	$b_{TS}(+1)$	+	$b_{CTS}(-1)$
8:	$y_8 = b_0$	+	$b_C(+1)$	+	$b_T(+1)$	+	$b_S(+1)$	+	$b_{CT}(+1)$	+	$b_{CS}(+1)$	+	$b_{TS}(+1)$	+	$b_{CTS}(+1)$

Our predictive model:

$$\hat{y} = b_0 + b_C\, x_C + b_T\, x_T + b_S\, x_S + b_{CT}\, x_C x_T + b_{CS}\, x_C x_S + b_{TS}\, x_T x_S + b_{CTS}\, x_C x_T x_S$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ 1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{pmatrix} \begin{pmatrix} b_0 \\ b_C \\ b_T \\ b_S \\ b_{CT} \\ b_{CS} \\ b_{TS} \\ b_{CTS} \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$


Safari File Edit View History Bookmarks Develop Window Help
r-fiddle.org
R-Fiddle
R-Fiddle
Save
Embed
Share
Install the R-Fiddle Chrome App
Like 266
C <- T <- S <- c(-1, +1)
design <- expand.grid(C=C, T=T, S=S)
C <- design$C
T <- design$T
S <- design$S
y <- c(5, 30, 6, 33, 4, 3, 5, 4)
# Avoid using this form - it is error prone:
# water <- lm(y ~ C + T + S + C*T + C*S + S*T + C*T*S)
# Rather use this form:
water <- lm(y ~ C*T*S)
summary(water)
library(pid)
paretoPlot(water)
Remember, you can always use the web-based version of R.
Go to http://yint.org/3C to run the code for this video.
Click here: http://yint.org/3C
Graphs
Run Code
>