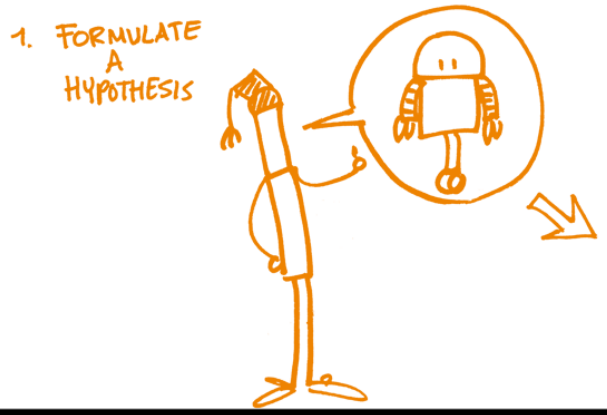
UNIVERSIDAD SIMÓN BOLÍVAR

ANGELA DI SERIO

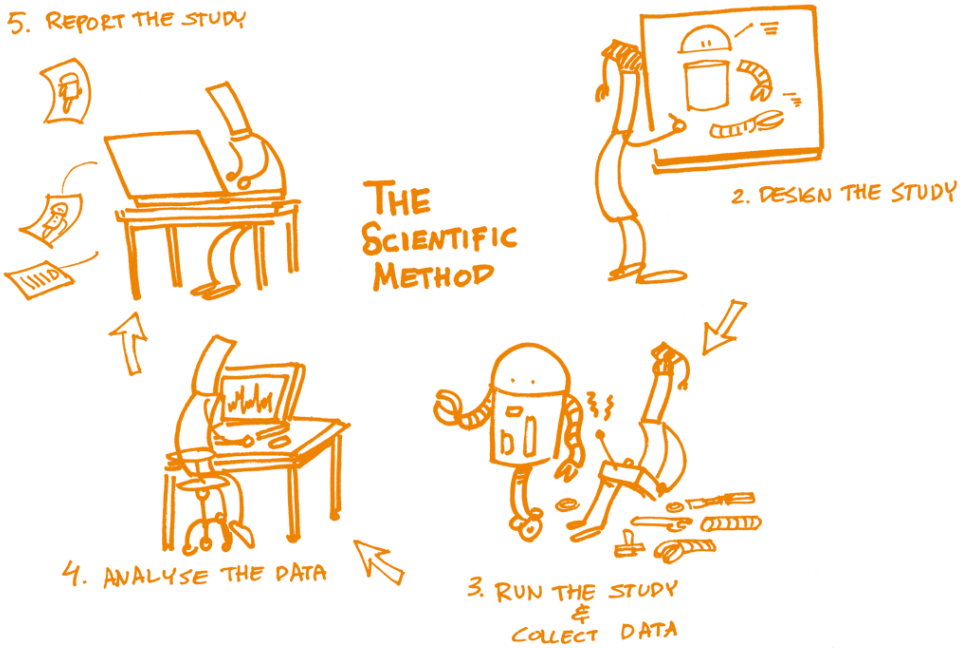
AGENDA

ASPECTOS A TRATAR

1. MOTIVACIÓN
2. DEFINICIÓN
3. INVESTIGACIÓN REPRODUCIBLE
4. BENEFICIOS
5. ¿QUÉ PODEMOS HACER?
6. ALTERNATIVA



MOTIVACIÓN 1



https://book.fosteropenscience.eu/es/02OpenScienceBasics/04ReproducibleResearchAndDataAnalysis.html

¿Por qué INVESTIGACIÓN REPRODUCIBLE?

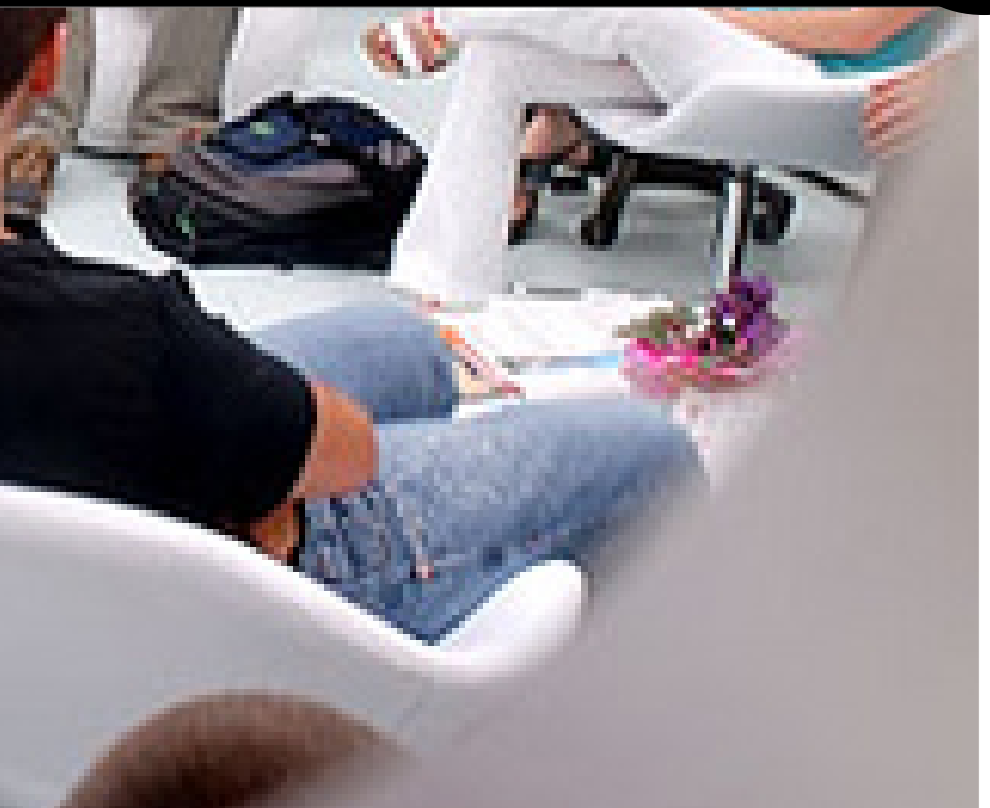
1. Motivación

¿POR QUÉ ES IMPORTANTE LA INVESTIGACIÓN REPRODUCIBLE?

Una variedad de estudios recientes, principalmente en el campo biomédico, han revelado que un número grande de resultados de investigación encontrados en la literatura fallan en reproducibilidad, debido a métodos experimentales descuidados, análisis estadísticos defectuosos o, en casos raros, fraude [1]

1 ACM - Artifact Review and Badging (Abril 2018)

DEFINICIÓN 2



DEBATE ACADÉMICO

Existe confusión entre los científicos sobre el significado de REPRODUCIBLE y REPLICABLE

REPRODUCIBILIDAD [2]

Capacidad de un investigador para duplicar los resultados de un estudio anterior utilizando los **mismos materiales** que utilizó el investigador original

REPLICABILIDAD

Capacidad de un investigador para duplicar los resultados de un estudio anterior si se siguen los **mismos procedimientos** pero se recopilan **datos nuevos**

2 Goodman, Fanelli y Ioannidis (2016) What does research reproducibility mean?

REPRODUCIBILIDAD [3]

La medición se puede obtener con la precisión indicada por un equipo diferente, un sistema de medición diferente, en una ubicación diferente en varias pruebas

En experimentos computacionales significa que un grupo independiente puede obtener el **mismo resultado utilizando artefactos** que desarrolla de forma completamente **independiente**

REPLICABILIDAD

La medición puede obtenerse con la precisión indicada por un equipo diferente utilizando el mismo procedimiento de medición, el mismo sistema de medición, en las mismas condiciones operativas, en la misma ubicación o en una ubicación diferente en múltiples pruebas

En experimentos computacionales, esto significa que un grupo independiente puede obtener el **mismo resultado utilizando los artefactos del autor**

3 ACM

REPRODUCIBILIDAD [4]

Un estudio es reproducible si puedes tomar los **datos originales** y el **código utilizado** para analizar los datos y reproducir todos los hallazgos numéricos del estudio

Esto puede parecer inicialmente una tarea trivial, pero la experiencia ha demostrado que no siempre es fácil lograr este estándar aparentemente mínimo

REPLICABILIDAD

Es la acción de repetir un estudio completo independientemente del investigador original sin el uso de datos originales pero generalmente utilizando los mismos métodos

4 American Statistical Association

INVESTIGACIÓN REPRODUCIBLE 3

Posibilidad de evaluar críticamente la exactitud de las afirmaciones científicas y conclusiones extraídas por otros científicos

Se requiere de una descripción precisa del procedimiento experimental y del posterior análisis de los datos, así como una cuidadosa atención a las posibles fuentes de error

Entre los primeros en escribir sobre reproducibilidad de esta manera está el geofísico Jon Claerbout quien fue pionero en el uso de la frase **investigación reproducible** para describir cómo su grupo de investigación en sismología utilizó archivos informáticos para permitir la regeneración eficiente de las figuras y tablas en tesis y publicaciones

Claerbout sentía frustación al trabajar con nuevos estudiantes que tenían grandes dificultades para duplicar la investigación de estudiantes anteriores

Hacer avances adicionales significaba pasar meses, o incluso años, tratando de reproducir avances anteriores

REPRODUCCIÓN

Cálculo de resultados científicos cuantitativos por **científicos independientes** que usan los **conjuntos de datos y métodos originales**

BENEFICIOS 4

4. Beneficios

1. Este valor debe ocultar algo

2. No indican el intervalo de confianza, me pregunto ¿será realmente significativo este valor?

3. ¿Puede ser esto cierto? Seguro removieron algunos valores

4. ¿Por qué muestran este gráfico en escala logarítmica ? ¿Cómo luce de la otra forma?

5. Solo muestran un subconjunto de los datos, ¿cómo luce el resto?

6. Si sólo pudiera acceder a los datos para verificar el p-valor de la prueba ...

FRUSTACIÓN COMO **REVISOR O LECTOR** DE ARTÍCULOS

1. ¿Qué datos usé para generar esta figura?

2. Funcionba ayer, ¿qué sucede hoy? Y unos meses después nos preguntaremos ¿por qué hice esto?

3. Uno de los revisores propone cambiar esta figura, ¿cómo la generé en su momento?

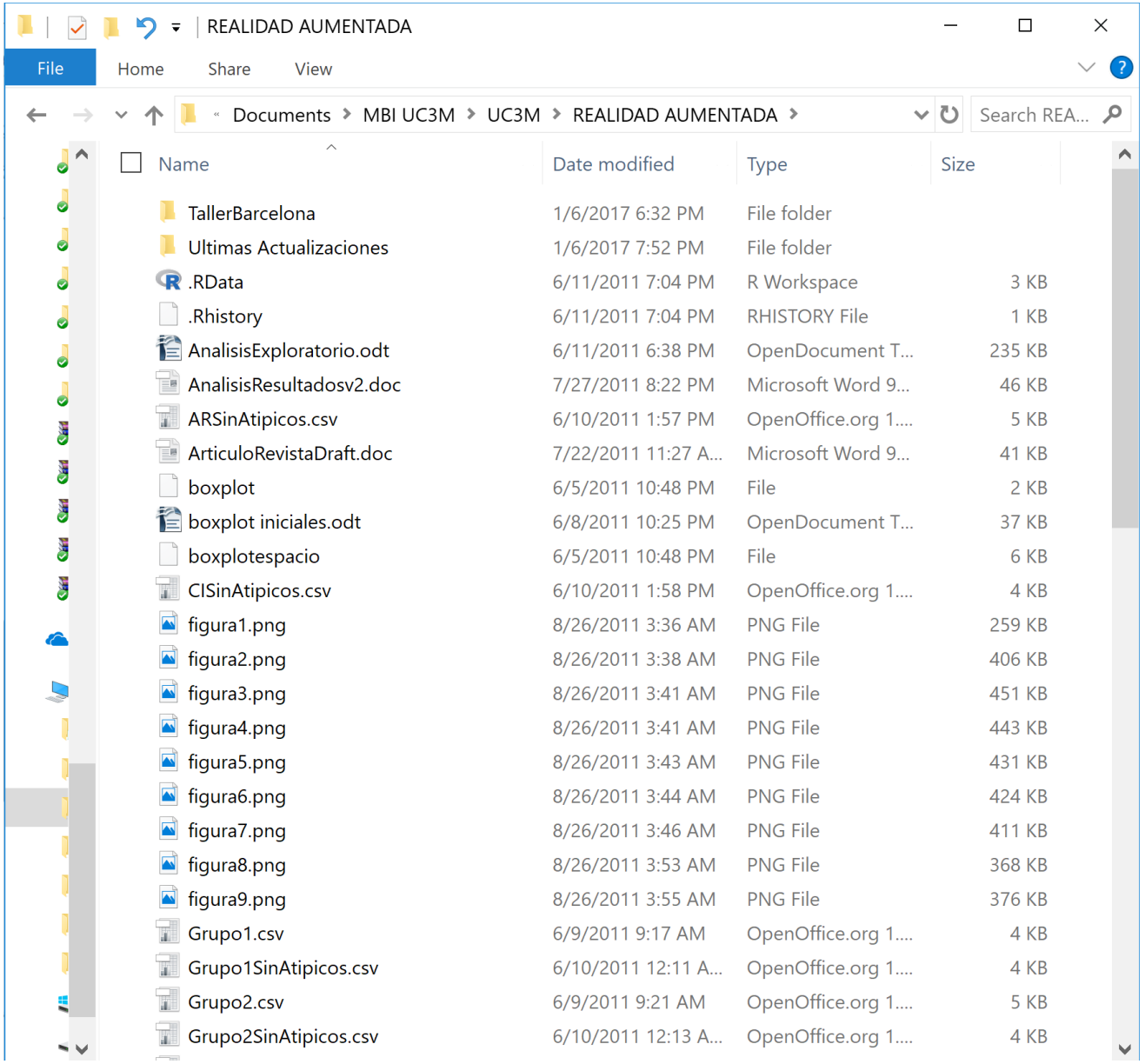
4. En la revisión me piden que agregue el intervalo de confianza y el efecto del tamaño, ¿qué datos usé?

5. Creo estar usando los mismos parámetros pero obtengo resultados diferentes

6. El nuevo tesista quiere comparar sus resultados con los resultados de otro, ¿dónde están los datos?

FRUSTACIÓN COMO **AUTOR** DE ARTÍCULOS

REALIDAD



4. Beneficios

Relevance

```
rel.tradicional<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Traditional Method"]
rel.tradicional
rel.AR<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Augmented Reality"]
rel.AR
var.test(rel.tradicional, rel.AR)
t.test(rel.tradicional, rel.AR, var.equal = F)
t.test(rel.tradicional, rel.AR, var.equal = T)
```

```
> rel.tradicional<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Traditional Method"]
> rel.tradicional
 [1] 2.88 3.33 3.33 2.67 3.75 3.22 3.33 3.63 3.22 3.00 2.78 3.44 4.13 2.89 3.00 2.22 3.75 4.00 3.11 3.44 3.56 2.78 2.67 3.33 3.56 3.78 4.22 3.00 3.50
[30] 3.22 4.33 3.25 2.33 3.11 3.44 3.33 3.78 2.44 3.00 3.89 3.22 3.78 2.89 3.44 2.67 2.33 3.44 3.00 3.56 3.44 2.67 3.33 3.89 3.22 4.00 3.44 3.56 2.89
[59] 3.56
> rel.AR<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Augmented Reality"]
> rel.AR
 [1] 3.89 3.11 3.56 3.33 3.56 3.33 2.78 3.67 3.11 3.33 3.33 3.00 3.44 4.00 3.78 4.22 2.67 3.89 3.33 3.78 3.71 3.00 3.33 4.56 4.22 4.00 3.11 3.56 3.22
[30] 4.25 2.38 3.78 3.44 4.22 3.67 3.56 3.33 3.22 4.44 3.56 3.67 3.11 3.50 4.67 3.56 2.78 4.00 3.78 3.11 2.56 2.63 3.67 3.33 2.67 3.33 4.33 3.22 4.33
[59] 4.44 3.44 3.13 4.33

> var.test(rel.tradicional, rel.AR)

        F test to compare two variances

data:  rel.tradicional and rel.AR
F = 0.8072, num df = 58, denom df = 61, p-value = 0.413
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4839619 1.3508779
sample estimates:
ratio of variances
         0.8072067
```

p-grande no se puede rechazar la hipotesis nula y se concluye que las varianzas son iguales
Aplicamos t-test asumiendo que las varianzas son iguales

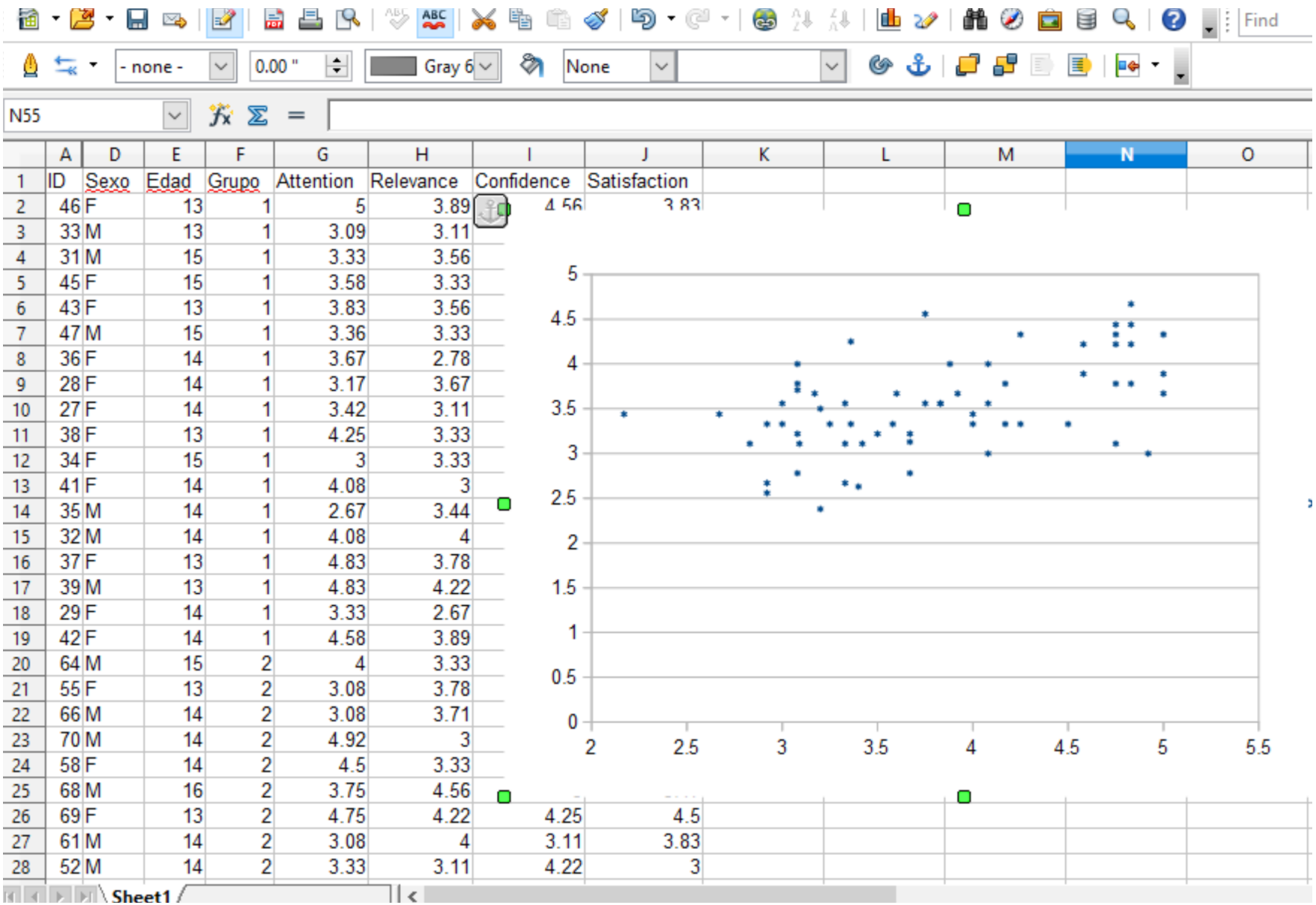
```
> t.test(rel.tradicional, rel.AR, var.equal = T)

        Two Sample t-test

data:  rel.tradicional and rel.AR
t = -2.7128, df = 119, p-value = 0.007662
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.43044524 -0.06720375
sample estimates:
mean of x mean of y
 3.287627  3.536452
```


Ayuda Cognitiva (Columna 4)

```
> t.test(Flujo[Flujo[,2]=="RA",4],Flujo[Flujo[,2]=="Web",4],paired=F)

        Welch Two Sample t-test

data:  Flujo[Flujo[, 2] == "RA", 4] and Flujo[Flujo[, 2] == "Web", 4]
t = 3.5828, df = 34.413, p-value = 0.001039
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2151415 0.7785093
sample estimates:
mean of x mean of y
 3.635714  3.138889

> wilcox.test(Flujo[Flujo[,2]=="RA",4],Flujo[Flujo[,2]=="Web",4],paired=F)

        Wilcoxon rank sum test with continuity correction

data:  Flujo[Flujo[, 2] == "RA", 4] and Flujo[Flujo[, 2] == "Web", 4]
W = 303.5, p-value = 0.001198
alternative hypothesis: true location shift is not equal to 0
```

Hay diferencia estadísticamente significativa para rechazar la hipótesis de igualdad de medias entre RA y Web

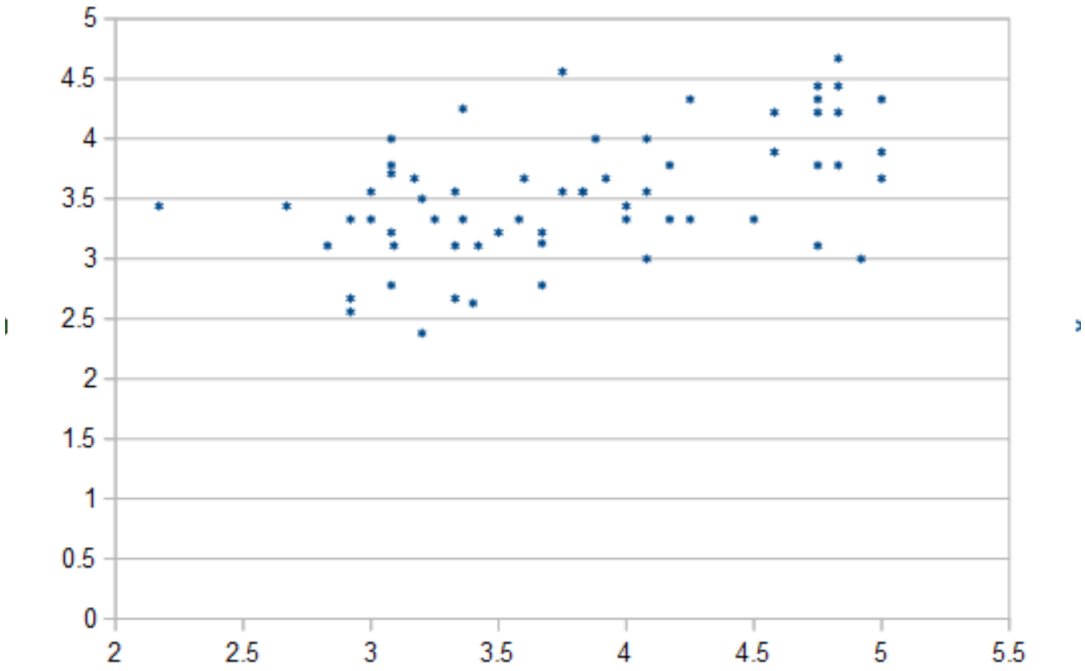
	RA	Web
Media	3.635714	3.138889
Desv	0.522920	0.334275
Median	3.67	3.085



	A	D	E	F	G	H	I	J
1	ID	Sexo	Edad	Grupo	Attention	Relevance	Confidence	Satisfaction
2	46	F	13	1	5	3.89	4.56	3.83
3	33	M	13	1	3.09	3.11		
4	31	M	15	1	3.33	3.56		
5	45	F	15	1	3.58	3.33		
6	43	F	13	1	3.83	3.56		
7	47	M	15	1	3.36	3.33		
8	36	F	14	1	3.67	2.78		
9	28	F	14	1	3.17	3.67		
10	27	F	14	1	3.42	3.11		
11	38	F	13	1	4.25	3.33		
12	34	F	15	1	3	3.33		
13	41	F	14	1	4.08	3		
14	35	M	14	1	2.67	3.44		
15	32	M	14	1	4.08	4		
16	37	F	13	1	4.83	3.78		
17	39	M	13	1	4.83	4.22		
18	29	F	14	1	3.33	2.67		
19	42	F	14	1	4.58	3.89		
20	64	M	15	2	4	3.33		
21	55	F	13	2	3.08	3.78		
22	66	M	14	2	3.08	3.71		
23	70	M	14	2	4.92	3		
24	58	F	14	2	4.5	3.33		
25	68	M	16	2	3.75	4.56		
26	69	F	13	2	4.75	4.22	4.25	4.5
27	61	M	14	2	3.08	4	3.11	3.83
28	52	M	14	2	3.33	3.11	4.22	3

4. Beneficios

QUÉ SUCEDE CUANDO COPIAMOS Y PEGAMOS ?



A medida que vamos trabajando con datos, nos vamos dando cuenta de lo poco convincentes que son algunos de los artículos que hemos leído o revisado y que muy probablemente hemos cometido errores en términos de metodología experimental o que nuestro trabajo no es fácilmente reproducible

1. Permite acelerar nuestra investigación al trabajar ordenadamente

2. Permite basar más fácilmente nuestro trabajo en trabajos previos

3. Facilita que otros investigadores basen sus trabajos en el nuestro y por lo tanto obtener más citas

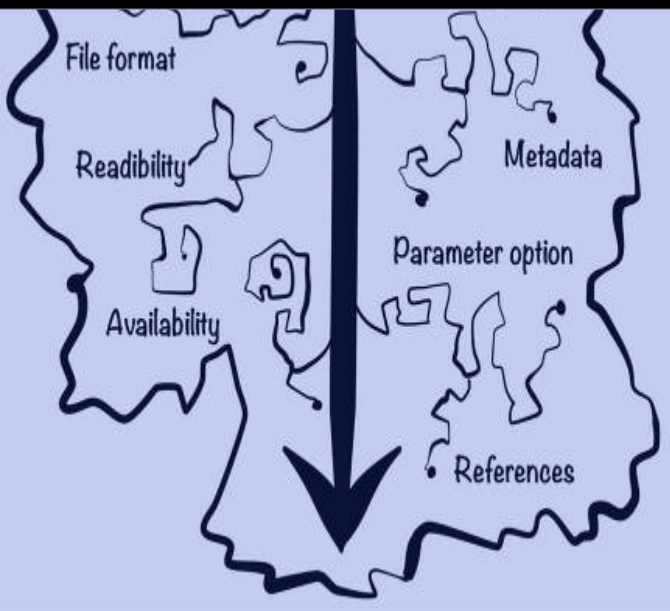
Y ...

¿CÓMO NOS BENEFICIA?

MÁS TEMPRANO O MÁS TARDE ...

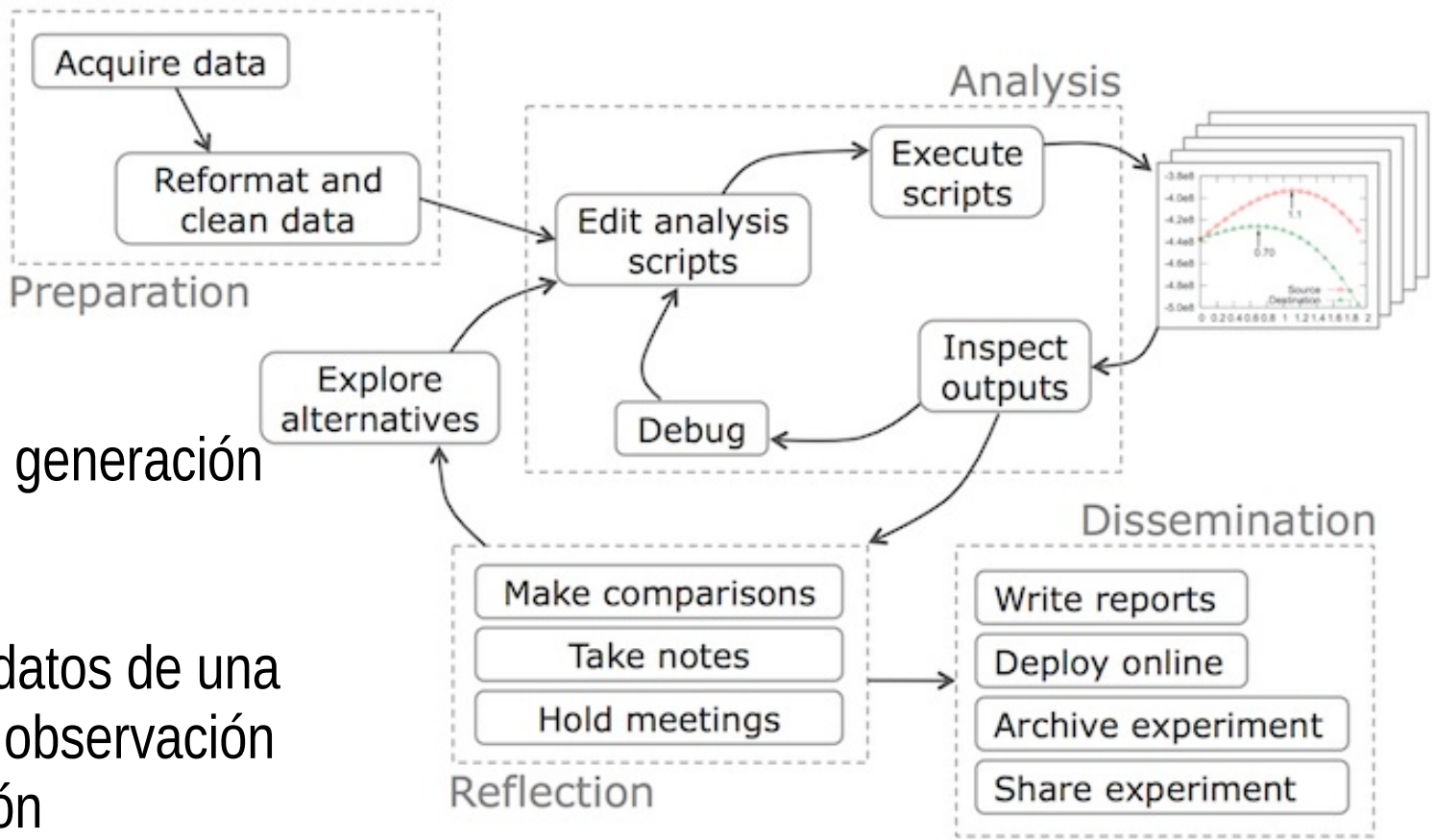
Nos veremos forzados a que nuestras publicaciones vayan acompañadas por los datos recogidos, los algoritmos utilizados, especificación paso a paso de las transformaciones, análisis, generación de gráficos y todo lo necesario para reproducir o replicar nuestros resultados

¿QUÉ PODEMOS HACER ? 5



Guillaume Dumas

WORKFLOW

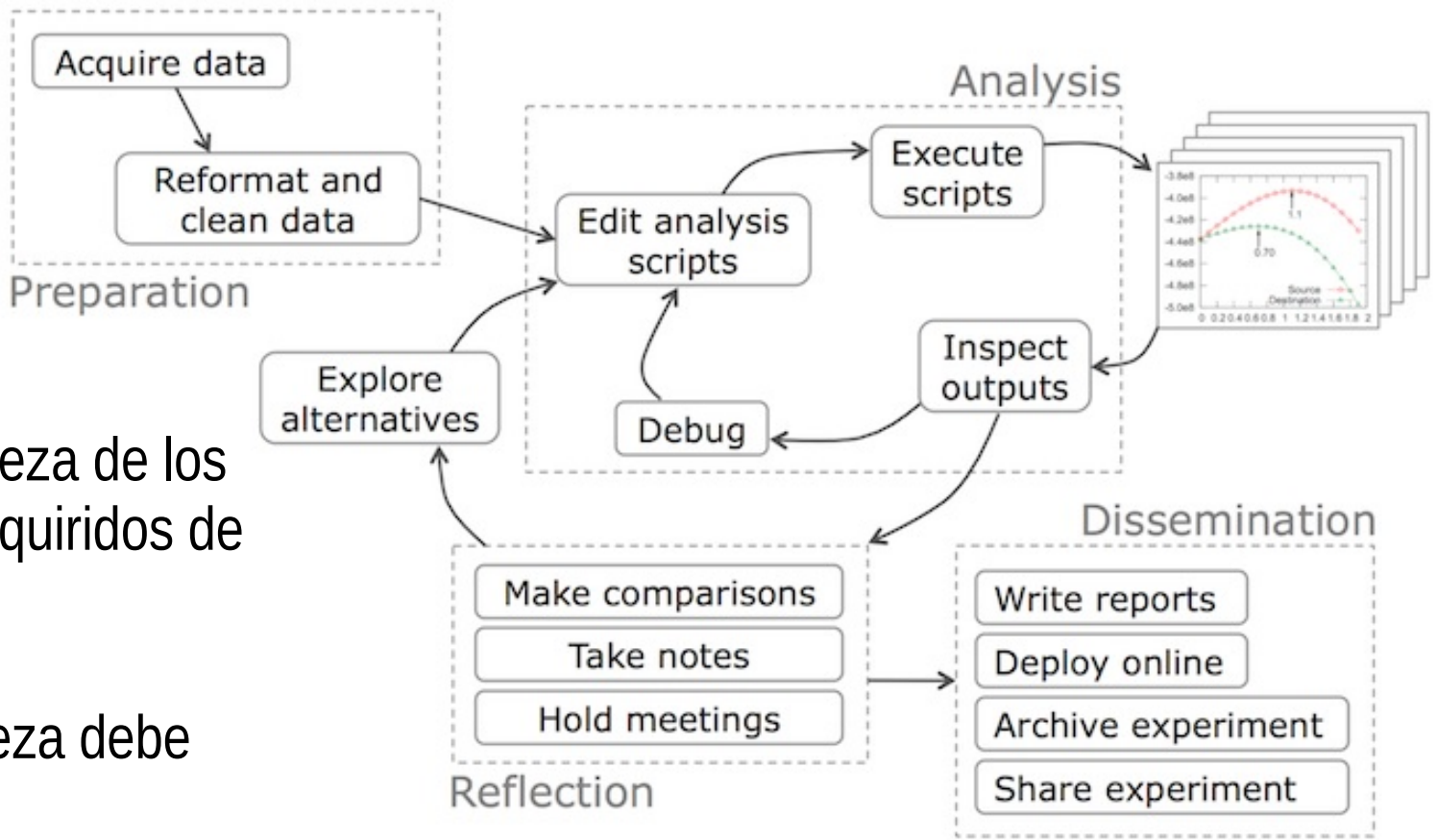


PREPARACIÓN

Adquisición, entrada o generación de los datos

Consiste en recopilar datos de una fuente primaria, como observación de campo, investigación experimental, cuestionarios o tests

WORKFLOW



PREPARACIÓN

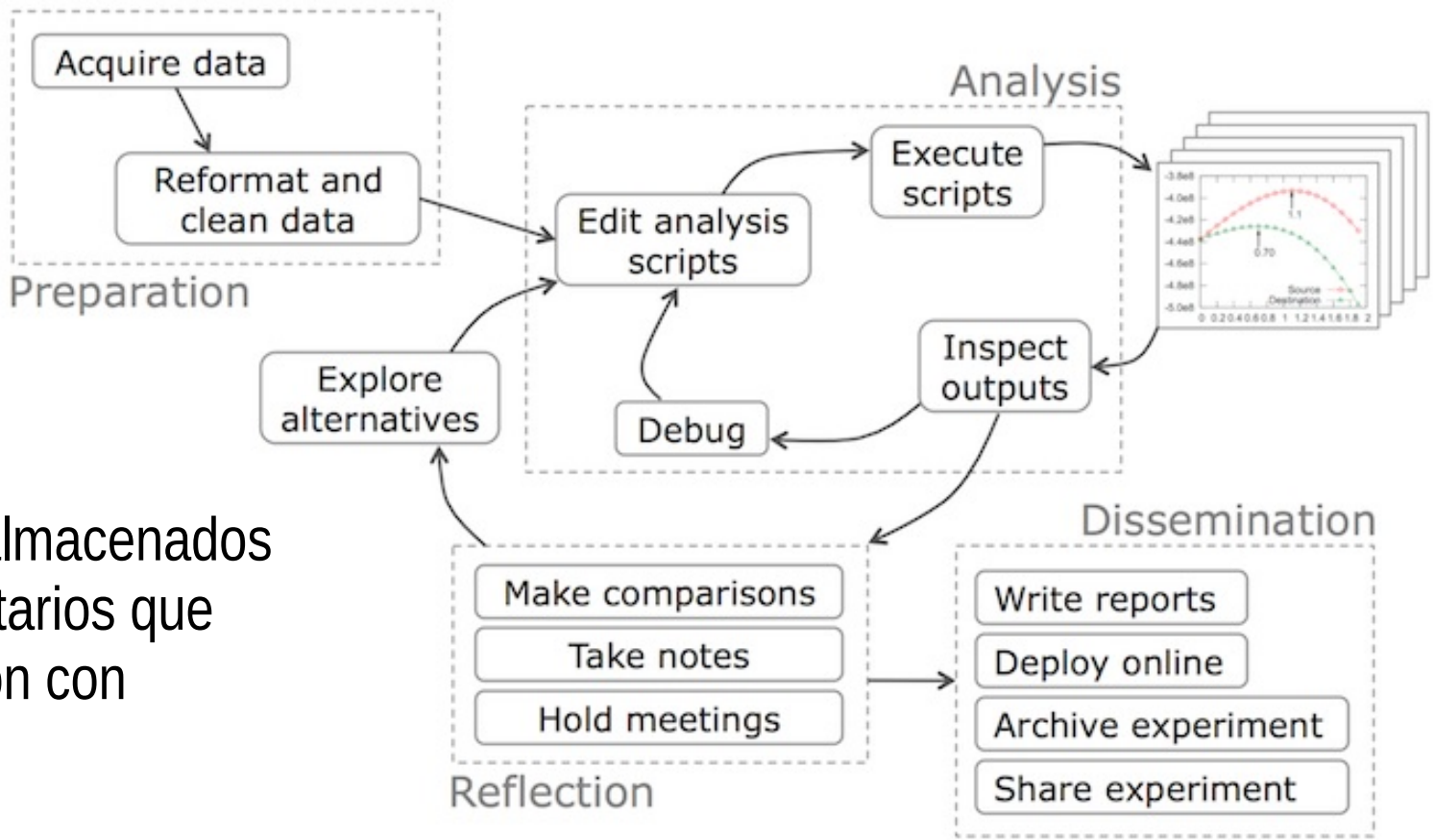
Procesamiento o limpieza de los datos producidos o adquiridos de fuentes externas

Este proceso de limpieza debe quedar explícitamente documentado incluyendo las decisiones tomadas

WORKFLOW

PREPARACIÓN

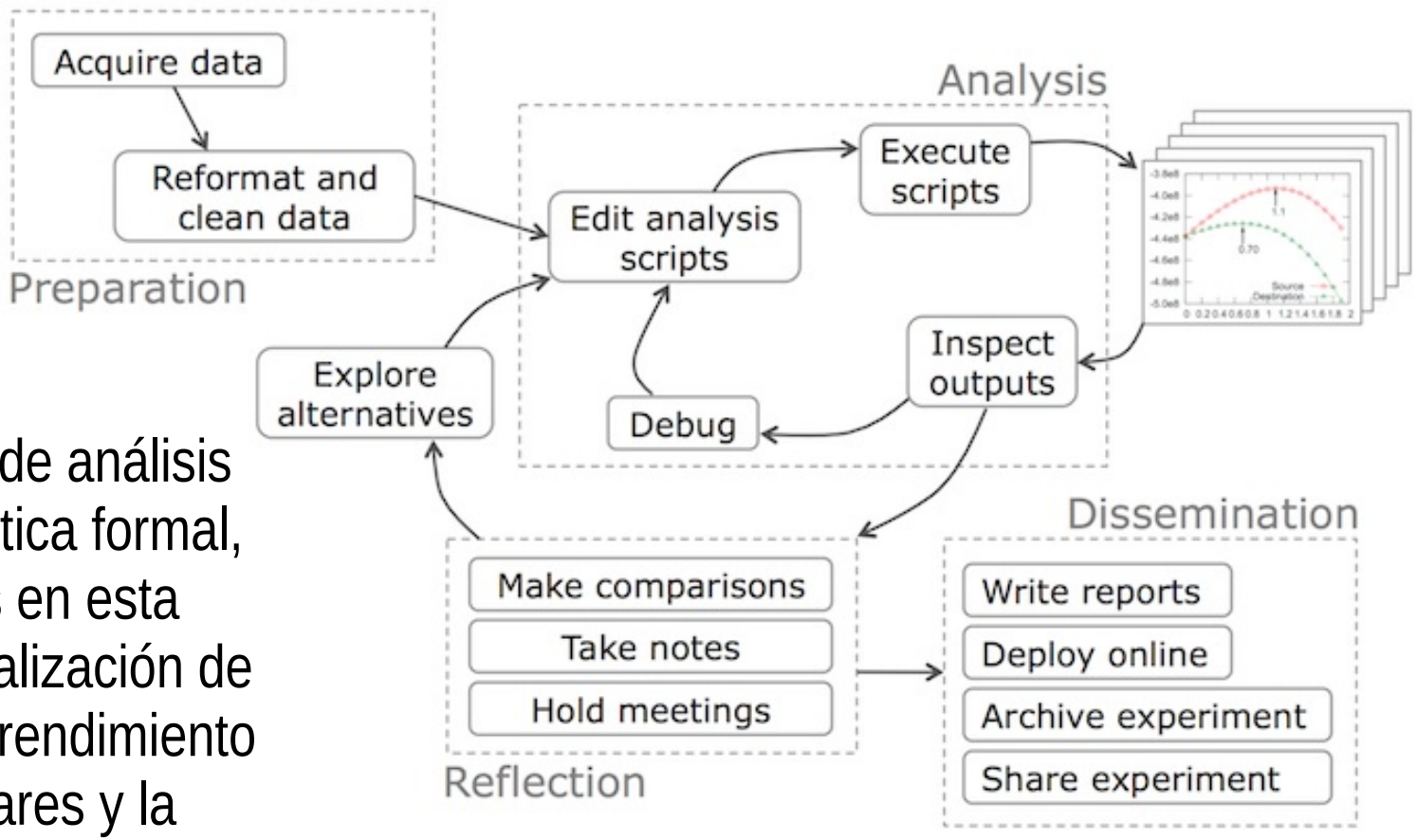
Los datos deben ser almacenados en formatos no propietarios que permitan su divulgación con facilidad



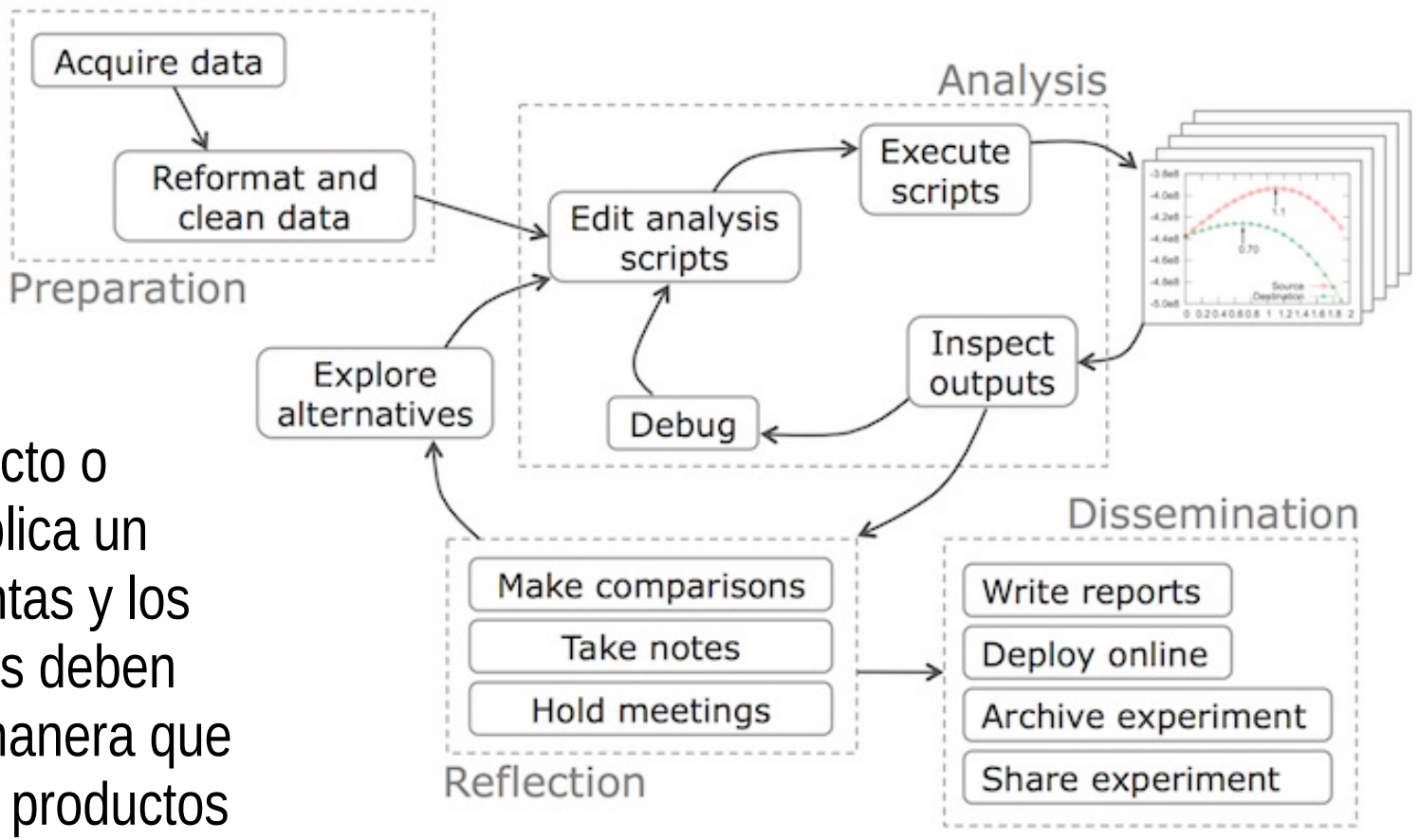
WORKFLOW

ANÁLISIS

La forma más común de análisis de datos es la estadística formal, pero otras actividades en esta etapa incluyen la visualización de datos, evaluación del rendimiento de algoritmos particulares y la extensión de los datos para abordar una hipótesis o extraer una conclusión científica

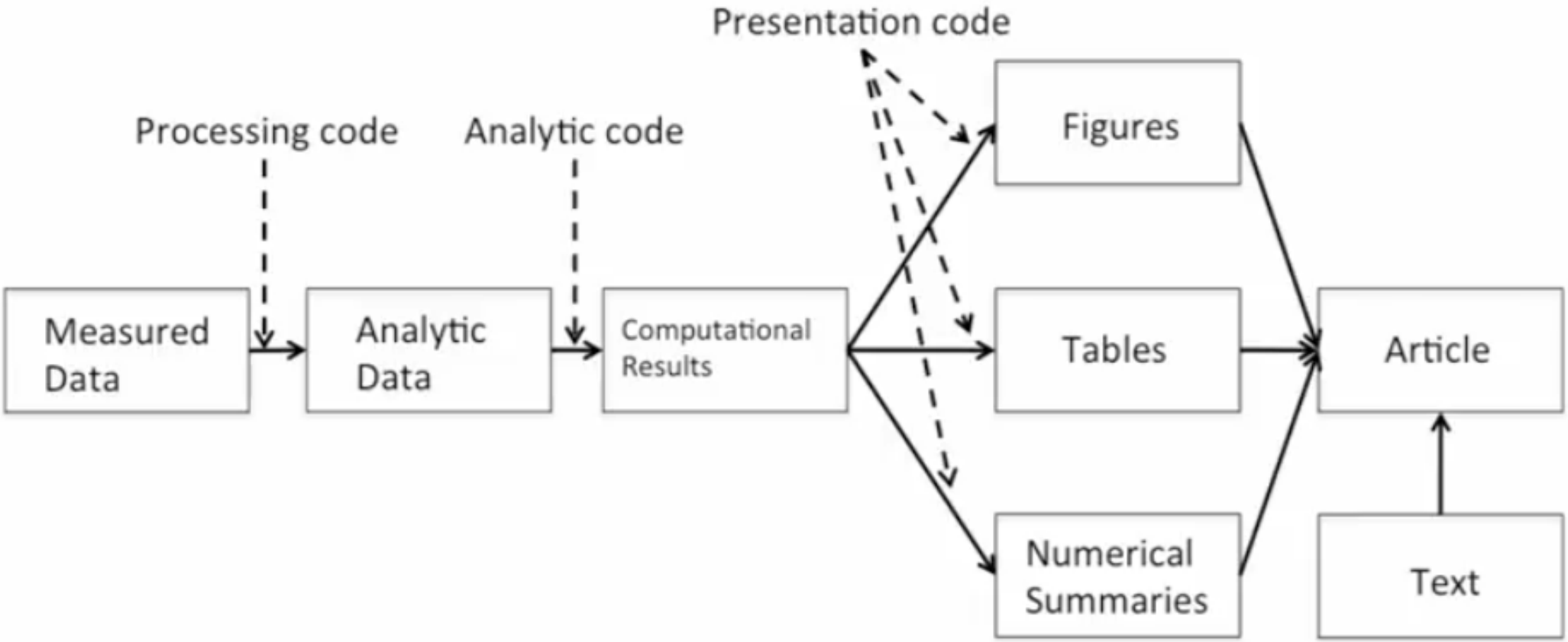


WORKFLOW



DIVULGACIÓN

Al completar un proyecto o investigación y se publica un artículo, las herramientas y los datos computacionales deben conservarse de una manera que permita reproducir los productos finales (figuras, tablas, valores de error) y que permita entender los métodos utilizados y su implementación



5.¿Qué podemos hacer?

Roger Peng. Report Writing for Data Science in R
https://leanpub.com/reportwriting

LITERATE PROGRAMMING

Introducido por Donald Knuth a principios de los años 80

Un programa tradicional consiste en un archivo de texto que contiene el código del programa y dispersos entre el código hay comentarios que describen las distintas partes del mismo

En LITERATE PROGRAMMING se invierte el énfasis

LITERATE PROGRAMMING

Permite escribir documentación que contiene código

El **programa** se convierte principalmente en un documento dirigido a los seres humanos

LITERATE PROGRAMMING

=

LEGIBLE POR LOS HUMANOS

+

LEGIBLE POR LOS ORDENADORES

ALTERNATIVA 6

R MARKDOWN

Herramienta disponible en el ambiente de R que nos permite construir y compartir el flujo de trabajo utilizado para llegar a los resultados y conclusiones de nuestro estudio

R MARKDOWN

Se basa en el lenguaje de marcado ligero conocido como MARKDOWN diseñado para ser legible por el humano y flexible en su salida

This is **bold** and _ italicizied _ text
This is **bold** and *italicizied* text

Entonces que es R Markdown ?

R MARKDOWN

=

MARKDOWN

+

Código en R

R MARKDOWN

1. Realmente soporta muchos lenguajes de programación y formatos de salida
2. R es el motor de lenguaje primario
3. R ha sido comúnmente utilizado para "envolver" o llamar a otros lenguajes de programación
4. Inicialmente Fortran y C, pero se ha expandido para incluir C++, Python, JavaScript, SQL, y más

R MARKDOWN

=

MARKDOWN

+

Código en R
(Fortran, C, C++, Python, Javascript, y otros)

ESTRUCTURA

Encabezado YAML

Cuerpo del Documento

ESTRUCTURA

Encabezado YAML

Cuerpo del Documento

YAML Ain't Markup Language

Utiliza la sintáxis de YAML para especificar información sobre el documento a generar

ESTRUCTURA

Encabezado YAML

```
---
title: "R Markdown - Parameterized"
output: html_document
params:
  start_date: !r Sys.Date() - 10
  end_date: !r Sys.Date()
---
```

Cuerpo del Documento

ESTRUCTURA

Encabezado YAML

Cuerpo del Documento

```
---
title: "Reproducible Research"
author: "Angela Di Serio"
date: "`r format(Sys.time(), '%d %B, %Y')`"
output: pdf_document
params:
  input_data: "datos.rds"
---
```

Algunas opciones de salida:

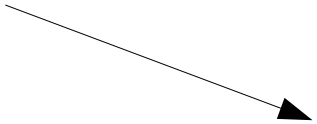
- HTML
- PDF
- Microsoft Word
- Shiny (interactivo)

Etc.

ESTRUCTURA

Encabezado YAML

Cuerpo del Documento



- Texto en formato Markdown
- Líneas de código
- Chunks o trozos de código
- Items embebidos como imágenes, etc

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
1  ---
2  title: "Effect of Vitamin C dose and supplement type on tooth growth"
3  author: "Angela Di Serio"
4  date: "October 2015"
5  output: pdf_document
6
7
8  ---
9
10
11 #Synopsis
12 In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg)
    on guinea pigs tooth growth  using two delivery methods: orange juice (OJ) or ascorbic acid
   (VC).
13 Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and
   orange juice) revealed a potential relationship between tooth length and supplement
   concentration. Statistical tests performed confirmed this observation. A higher dose level
   of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the
   delivery methods, we can conclude that the orange juice produces longer teeth for dose
   levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no
   difference in tooth lenght.
14
15 #Getting Data
16 The data used for this study were taken from the R datasets package. It consists of a data
   frame with 60 observations on 3 variables:
17
18 ``` {r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}
19
20 list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
21 new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
22 if(length(new.packages)) install.packages(new.packages)
23
24 library(datasets)
25 library(pander)
26 library(ggplot2)
27 library(knitr)
28 library(sjmisc)
```

32:1 | Getting Data | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 1 / 7 Find:

Effect of Vitamin C dose and supplement type on tooth growth

Angela Di Serio

October 2015

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

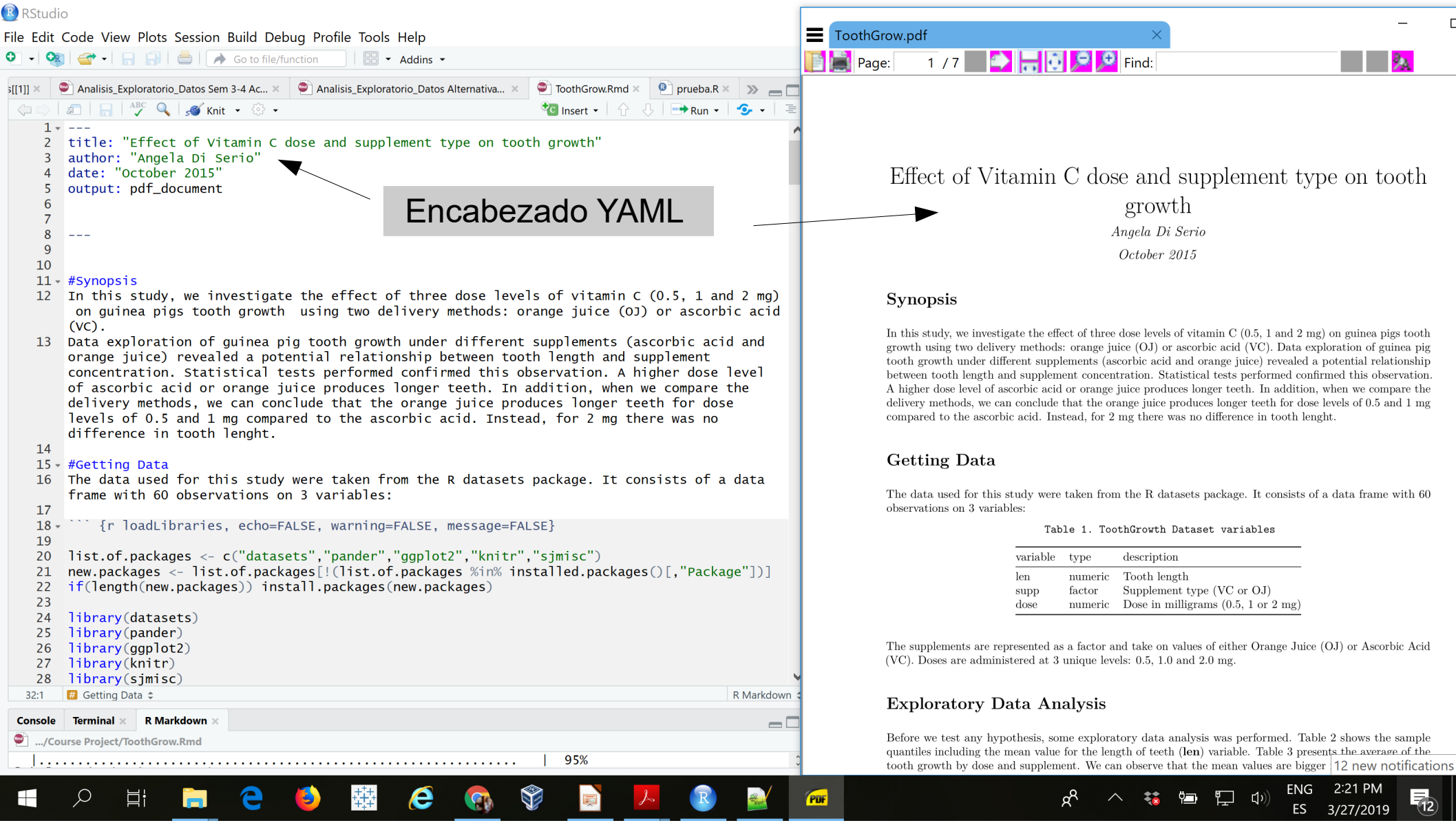
Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger

12 new notifications

ENG ES 2:21 PM 3/27/2019

6. Alternativa



6. Alternativa

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
1  ---
2  title: "Effect of Vitamin C dose and supplement type on tooth growth"
3  author: "Angela Di Serio"
4  date: "October 2015"
5  output: pdf_document
6
7
8  ---
9
10
11 #Synopsis
12 In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg)
    on guinea pigs tooth growth  using two delivery methods: orange juice (OJ) or ascorbic acid
   (VC).
13 Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and
   orange juice) revealed a potential relationship between tooth length and supplement
   concentration. Statistical tests performed confirmed this observation. A higher dose level
   of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the
   delivery methods, we can conclude that the orange juice produces longer teeth for dose
   levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no
   difference in tooth lenght.
14
15 #Getting Data
16 The data used for this study were taken from the R datasets package. It consists of a data
   frame with 60 observations on 3 variables:
17
18 ``` {r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}
19
20 list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
21 new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
22 if(length(new.packages)) install.packages(new.packages)
23
24 library(datasets)
25 library(pander)
26 library(ggplot2)
27 library(knitr)
28 library(sjmisc)
```

Texto en Markdown

32:1 | Getting Data | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 1 / 7 | Find:

Effect of Vitamin C dose and supplement type on tooth growth

Angela Di Serio

October 2015

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger

12 new notifications

ENG ES 2:21 PM 3/27/2019

6. Alternativa

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
1  ---
2  title: "Effect of Vitamin C dose and supplement type on tooth growth"
3  author: "Angela Di Serio"
4  date: "October 2015"
5  output: pdf_document
6
7
8  ---
9
10
11 #Synopsis
12 In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg)
    on guinea pigs tooth growth  using two delivery methods: orange juice (OJ) or ascorbic acid
   (VC).
13 Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and
   orange juice) revealed a potential relationship between tooth length and supplement
   concentration. Statistical tests performed confirmed this observation. A higher dose level
   of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the
   delivery methods, we can conclude that the orange juice produces longer teeth for dose
   levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no
   difference in tooth lenght.
14
15 #Getting Data
16 The data used for this study were taken from the R datasets package. It consists of a data
   frame with 60 observations on 3 variables:
17
18 ``` {r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}
19
20 list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
21 new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
22 if(length(new.packages)) install.packages(new.packages)
23
24 library(datasets)
25 library(pander)
26 library(ggplot2)
27 library(knitr)
28 library(sjmisc)
```

Trozo de código

32:1 | Getting Data | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 1 / 7 Find:

Effect of Vitamin C dose and supplement type on tooth growth

Angela Di Serio

October 2015

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger

12 new notifications

ENG ES 2:21 PM 3/27/2019

6. Alternativa

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
27 library(knitr)
28 library(sjmisc)
29
30 ```
31
32
33                          Table 1. ToothGrowth Dataset variables
34
35 | variable | type    | description                          |
36 |----------|---------|--------------------------------------|
37 | len      | numeric | Tooth length                         |
38 | supp     | factor  | Supplement type (VC or OJ)           |
39 | dose     | numeric | Dose in milligrams (0.5, 1 or 2 mg)  |
40
41
42 The supplements are represented as a factor and take on values of either Orange Juice (OJ)
   or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.
43
44
45 ```{r getData, results="asis", echo=FALSE, eval=FALSE}
46 pandoc.table(head(ToothGrowth),"First 5 records from dataset TootGrowth")
47 ```
48
49
50 #Exploratory Data Analysis
51 Before we test any hypothesis, some exploratory data analysis was performed.
52 Table 2 shows the sample quantiles including the mean value for the length of teeth
   (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement.
   We can observe that the mean values are bigger for Vitamin C when the doses are 0.5 mg or 1
   mg. In the case of 2 mg,  the mean values are very similar either for VC or OJ.
53
54 ```{r eda, results="asis", echo = FALSE, warning=FALSE }
55 pandoc.table(summary(ToothGrowth$len),"Descriptive statistics for Tooth Growth")
56 pandoc.table(with(ToothGrowth, tapply(len, list(supp,dose), mean)),"Mean values of Tooth
   Growth by dose and supplement")
57
58 ```
```

58:1 | Chunk 3: eda | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 1 / 7 Find:

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger for Vitamin C when the doses are 0.5 mg or 1 mg. In the case of 2 mg, the mean values are very similar either for VC or OJ.

Table 2: Descriptive statistics for Tooth Growth

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.2	13.08	19.25	18.81	25.28	33.9

1

ENG ES 2:36 PM 3/27/2019 14

6. Alternativa

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
27 library(knitr)
28 library(sjmisc)
29
30 ```
31
32
33                          Table 1. ToothGrowth Dataset variables
34
35 | variable | type    | description                          |
36 |----------|---------|--------------------------------------|
37 | len      | numeric | Tooth length                         |
38 | supp     | factor  | Supplement type (VC or OJ)           |
39 | dose     | numeric | Dose in milligrams (0.5, 1 or 2 mg)  |
40
41
42 The supplements are represented as a factor and take on values of either Orange Juice (OJ)
   or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.
43
44
45 ``` {r getData, results="asis", echo=FALSE, eval=FALSE}
46 pandoc.table(head(ToothGrowth),"First 5 records from dataset TootGrowth")
47 ```
48
49
50 #Exploratory Data Analysis
51 Before we test any hypothesis, some exploratory data analysis was performed.
52 Table 2 shows the sample quantiles including the mean value for the length of teeth
   (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement.
   We can observe that the mean values are bigger for Vitamin C when the doses are 0.5 mg or 1
   mg. In the case of 2 mg,  the mean values are very similar either for VC or OJ.
53
54 ``` {r eda, results="asis", echo = FALSE, warning=FALSE }
55 pandoc.table(summary(ToothGrowth$len),"Descriptive statistics for Tooth Growth")
56 pandoc.table(with(ToothGrowth, tapply(len, list(supp,dose), mean)),"Mean values of Tooth
   Growth by dose and supplement")
57
58 ```
```

58:1 | Chunk 3: eda | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 1 / 7 Find:

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger for Vitamin C when the doses are 0.5 mg or 1 mg. In the case of 2 mg, the mean values are very similar either for VC or OJ.

Table 2: Descriptive statistics for Tooth Growth

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.2	13.08	19.25	18.81	25.28	33.9

1

ENG ES 2:36 PM 3/27/2019 14

6. Alternativa

Investigacion Reproducible 05-04-2019 Version Libre Office.odp - LibreOffice Impress

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

atorio_Datos Alternativa... | prueba.R | prueba_notebook.Rmd | ToothGrow.Rmd* | pruebaPython.Rmd | mydfs

Knit | Insert | Run

```
32
33  ```
34
35
36                          Table 1. ToothGrowth Dataset variables
37
38  | variable | type    | description                        |
39  |----------|---------|------------------------------------|
40  | len      | numeric | Tooth length                       |
41  | supp     | factor  | Supplement type (VC or OJ)         |
42  | dose     | numeric | Dose in milligrams (0.5, 1 or 2 mg) |
43
44
45  The supplements are represented as a factor and take on values of either Orange Juice (OJ)
    or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.
46
47
48  ```{r getData, results="asis", echo=FALSE, eval=FALSE}
49  pandoc.table(head(ToothGrowth),"First 5 records from dataset TootGrowth")
50  ```
51
52
53  #Exploratory Data Analysis
54  Before we test any hypothesis, some exploratory data analysis was performed.
55  Table 2 shows the sample quantiles including the mean value for the length of teeth
    (**len**) variable. Table 3 presents the average of the tooth growth by dose and
    supplement. We can observe that the mean values are bigger for Vitamin C when the doses
    are 0.5 mg or 1 mg. In the case of 2 mg,  the mean values are very similar either for VC
    or OJ.
56
57  ```{r eda, results="asis", echo = FALSE, warning=FALSE }
58  pandoc.table(summary(ToothGrowth$len),"Descriptive statistics for Tooth Growth")
59
60  ```
61
62  A visual representation of the data helps us understand the relationship between doses and
    supplements. Figures 1 and 2 present boxplots of the data  with different arrangements.
63
64  ```{r Boxplot1, echo=FALSE }
```

59:1 | Chunk 3: eda | R Markdown

Console

ToothGrow.pdf

Page: 1 / 7 | Find:

October 2015

Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

Getting Data

The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

variable	type	description
len	numeric	Tooth length
supp	factor	Supplement type (VC or OJ)
dose	numeric	Dose in milligrams (0.5, 1 or 2 mg)

The supplements are represented as a factor and take on values of either Orange Juice (OJ) or Ascorbic Acid (VC). Doses are administered at 3 unique levels: 0.5, 1.0 and 2.0 mg.

Exploratory Data Analysis

Before we test any hypothesis, some exploratory data analysis was performed. Table 2 shows the sample quantiles including the mean value for the length of teeth (**len**) variable. Table 3 presents the average of the tooth growth by dose and supplement. We can observe that the mean values are bigger for Vitamin C when the doses are 0.5 mg or 1 mg. In the case of 2 mg, the mean values are very similar either for VC or OJ.

Table 2: Descriptive statistics for Tooth Growth

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.2	13.08	19.25	18.81	25.28	33.9

A visual representation of the data helps us understand the relationship between doses and supplements. Figures 1 and 2 present boxplots of the data with different arrangements.

ENG US 3:13 PM 3/28/2019

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
53
54  ```{r eda, results="asis", echo = FALSE, warning=FALSE }
55  pandoc.table(summary(ToothGrowth$len),"Descriptive statistics for Tooth Growth")
56  pandoc.table(with(ToothGrowth, tapply(len, list(supp,dose), mean)),"Mean values of Tooth
    Growth by dose and supplement")
57
58  ```
59
60  A visual representation of the data helps us understand the relationship between doses and
    supplements. Figures 1 and 2 present boxplots of the data  with different arrangements.
61
62  ```{r Boxplot1, echo=FALSE,}
63  ggplot(aes(x=as.factor(dose),y=len),data=ToothGrowth) +
64  geom_boxplot(aes(fill=supp)) +
65  facet_wrap(~supp)+
66  xlab("Dose")+
67  ggtitle("Figure 1. Effect of Vitamin C dose on tooth growth\n given type of supplement")+
68  theme(axis.text=element_text(size=10),axis.title=element_text(size=10),plot.title =
    element_text(size=11))
69  ```
70
71  For doses of 0.5 mg and 1.0 mg, orange juice appears to help tooth growth more than ascorbic
    acid. However in the case of a dose of 2.0mg, there does not appear to be one supplement
    that is better than the other. Also, for either supplement type, we observe a positive
    correlation between tooth growth and dose.
72
73  Data exploration of Guinea Pig tooth growth under different supplements (Vitamin C and
    Orange Juice) revealed a potential relationship of tooth length vs supplement concentration.
    In order to investigate this finding, different hypothesis testings were conducted.
74
75
76  ```{r Boxplot2, echo=FALSE}
77  ggplot(aes(x=supp,y=len),data=ToothGrowth) +
78  geom_boxplot(aes(fill=supp)) +
79  facet_wrap(~dose)+
80  xlab("Supplement")+
81  ggtitle("Figure 2. Effect of Vitamin C dose and supplement type\n on tooth growth in Guinea
```

58:1 | Chunk 3: eda | R Markdown

Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

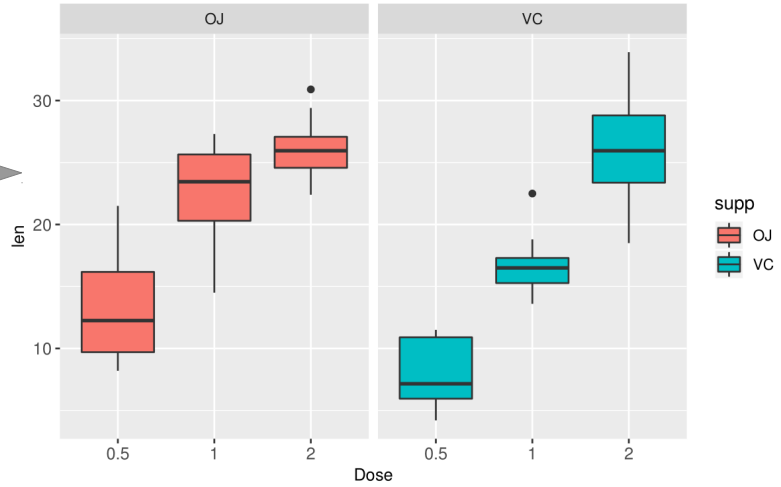
ToothGrow.pdf

Page: 2 / 7 Find:

	0.5	1	2
OJ	13.23	22.7	26.06
VC	7.98	16.77	26.14

A visual representation of the data helps us understand the relationship between doses and supplements. Figures 1 and 2 present boxplots of the data with different arrangements.

Figure 1. Effect of Vitamin C dose on tooth growth given type of supplement



For doses of 0.5 mg and 1.0 mg, orange juice appears to help tooth growth more than ascorbic acid. However in the case of a dose of 2.0mg, there does not appear to be one supplement that is better than the other. Also, for either supplement type, we observe a positive correlation between tooth growth and dose.

Data exploration of Guinea Pig tooth growth under different supplements (Vitamin C and Orange Juice) revealed a potential relationship of tooth length vs supplement concentration. In order to investigate this finding, different hypothesis testings were conducted.

ENG ES 2:57 PM 3/27/2019

3. Investigación Reproducible

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Analisis_Exploratorio_Datos Sem 3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R

Knit | Insert | Run

```
88  The purpose of this research study was to investigate the effect of three dose levels of
    vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods (orange
    juice or ascorbic acid).
89
90  The research questions were:
91
92  1. Is there any difference in tooth growth depending on which delivery method is used for
    each dose level?
93  2. Are there any differences in tooth growth depending on the dose level for each
    supplement?
94
95  #Results
96  ## Research Question 1
97  Is there any difference in tooth growth depending on which delivery method is used for each
    dose level?
98
99  t-tests were used to compare the two delivery methods (Orange Juice and Ascorbic Acid) for
    the different doses:
100
101 1. $H_o: \mu_{oj} == \mu_{vc}$, $H_a: \mu_{oj} <> \mu_{vc}$, dose level = 0.5
102
103 2. $H_o: \mu_{oj} == \mu_{vc}$, $H_a: \mu_{oj} <> \mu_{vc}$, dose level = 1
104
105 3. $H_o: \mu_{oj} == \mu_{vc}$, $H_a: \mu_{oj} <> \mu_{vc}$, dose level = 2
106
107 Since we performed multiple tests of significance, the Bonferroni correction procedure was
    applied to prevent a chance of making a type I error.  Homogeneity of variances was checked
    using Levene's test confirming that all the variances were equal. Table 4 shows t-test
    results.
108
109
110 ```{r Test1, echo=FALSE}
111
112 aux1 <- t.test(ToothGrowth[which(ToothGrowth$dose==0.5 & ToothGrowth$supp=='OJ'),]$len,
113 +        ToothGrowth[which(ToothGrowth$dose==0.5 & ToothGrowth$supp=='VC'),]$len,
    paired=FALSE,var.equal = TRUE)
114 aux2 <- t.test(ToothGrowth[which(ToothGrowth$dose==1 & ToothGrowth$supp=='OJ'),]$len,
```

58:1 | Chunk 3: eda | R Markdown

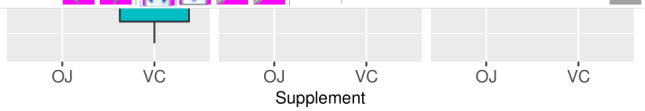
Console | Terminal | R Markdown

.../Course Project/ToothGrow.Rmd

|.....| 95%

ToothGrow.pdf

Page: 3 / 7 Find:



Research Questions

The purpose of this research study was to investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods (orange juice or ascorbic acid).

The research questions were:

1. Is there any difference in tooth growth depending on which delivery method is used for each dose level?
2. Are there any differences in tooth growth depending on the dose level for each supplement?

Results

Research Question 1

Is there any difference in tooth growth depending on which delivery method is used for each dose level?

t-tests were used to compare the two delivery methods (Orange Juice and Ascorbic Acid) for the different doses:

1. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 0.5
2. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 1
3. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 2

3

Since we performed multiple tests of significance, the Bonferroni correction procedure was applied to prevent a chance of making a type I error. Homogeneity of variances was checked using Levene's test confirming that all the variances were equal. Table 4 shows t-test results.

Table 4. t-tests results (OJ vs VC)

ENG ES 2:59 PM 3/27/2019

6. Alternativa

~/Documents/Investigacion Reproducible/Proyecto R Reproducible - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

PruebaGradient.Rmd

Knit | Insert | Run

```
120
121 Since we performed multiple tests of significance, the Bonferroni correction procedure was
    applied to prevent a chance of making a type I error.  Homogeneity of variances was checked
    using Levene's test confirming that all the variances were equal. Table 4 shows t-test results.
122
123
124 ```{r Test1, echo=FALSE}
125
126 aux1 <- t.test(ToothGrowth[which(ToothGrowth$dose==0.5 & ToothGrowth$supp=='OJ'),]$len,
127 +        ToothGrowth[which(ToothGrowth$dose==0.5 & ToothGrowth$supp=='VC'),]$len,
    paired=FALSE,var.equal = TRUE)
128 aux2 <- t.test(ToothGrowth[which(ToothGrowth$dose==1 & ToothGrowth$supp=='OJ'),]$len,
129 +        ToothGrowth[which(ToothGrowth$dose==1 & ToothGrowth$supp=='VC'),]$len,
    paired=FALSE,var.equal = TRUE)
130 aux3 <- t.test(ToothGrowth[which(ToothGrowth$dose==2 & ToothGrowth$supp=='OJ'),]$len,
131 +        ToothGrowth[which(ToothGrowth$dose==2 & ToothGrowth$supp=='VC'),]$len,
    paired=FALSE,var.equal = TRUE)
132
133 p<-p.adjust(c(aux1$p.value,aux2$p.value,aux3$p.value),method="bonferroni")
134
135 ```
136
137                                   Table 4. t-tests results (OJ vs VC)
138
139 | Test Statistic | df | Confidence Interval | p-value |Adjusted p | Dose mg |
140 |---------------:|----|:-------------------:|---------|:-------:|:--------:|
141 | `r round(aux1$statistic,2)` | `r round(aux1$parameter,2)` | `r
    as.character(round(aux1$conf.int,2))` | `r round(aux1$p.value,3)` | `r round(p[1],3)` | 0.5 |
142 | `r round(aux2$statistic,2)` | `r round(aux2$parameter,2)` | `r
    as.character(round(aux2$conf.int,2))` | `r round(aux2$p.value,3)` | `r round(p[2],3)` | 1 |
143 | `r round(aux3$statistic,2)` | `r round(aux3$parameter,2)` | `r
    as.character(round(aux3$conf.int,2))` | `r round(aux3$p.value,3)` | `r round(p[3],3)` | 2 |
144
145
146 The results obtained when we compare the delivery methods using a unpaired-sample t test
    indicate that there is a statistically significant difference when the dose of Vitamin C is
    equal to 0.5 or 1 mg. The adjusted p-values for these two cases is less than 0.05 and therefore,
    we reject the null hyphotesis. Also, the confidence intervals for dose level 0.5 and 1 mg do not
    contain the zero value. For the 2 mg dose level, the adjusted p-value is greater than 0.05 and
```

213:75 Chunk 9: Test3 R Markdo

Console

PruebaGradient.pdf

Page: 4 / 8 Find:

Research Question 1

Is there any difference in tooth growth depending on which delivery method is used for each dose level?

t-tests were used to compare the two delivery methods (Orange Juice and Ascorbic Acid) for the different doses:

1. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 0.5
2. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 1
3. $H_o : \mu_{oj} == \mu_{vc}$, $H_a : \mu_{oj} <> \mu_{vc}$, dose level = 2

4

Since we performed multiple tests of significance, the Bonferroni correction procedure was applied to prevent a chance of making a type I error. Homogeneity of variances was checked using Levene's test confirming that all the variances were equal. Table 4 shows t-test results.

Table 4. t-tests results (OJ vs VC)

Test Statistic	df	Confidence Interval	p-value	Adjusted p	Dose mg
3.17	18	1.77, 8.73	0.005	0.016	0.5
4.03	18	2.84, 9.02	0.001	0.002	1
-0.05	18	-3.72, 3.56	0.964	1	2

The results obtained when we compare the delivery methods using a unpaired-sample t test indicate that there is a statistically significant difference when the dose of Vitamin C is equal to 0.5 or 1 mg. The adjusted p-values for these two cases is less than 0.05 and therefore, we reject the null hyphotesis. Also, the confidence intervals for dose level 0.5 and 1 mg do not contain the zero value. For the 2 mg dose level, the adjusted p-value is greater than 0.05 and we do not have evidence to reject the null hyphotesis. The confidence interval for this case contains the 0 value.

Research Question 2

Are there any differences in tooth growth depending on the dose level for each supplement?

Effect of different levels of Vitamin C using orange juice as delivery method

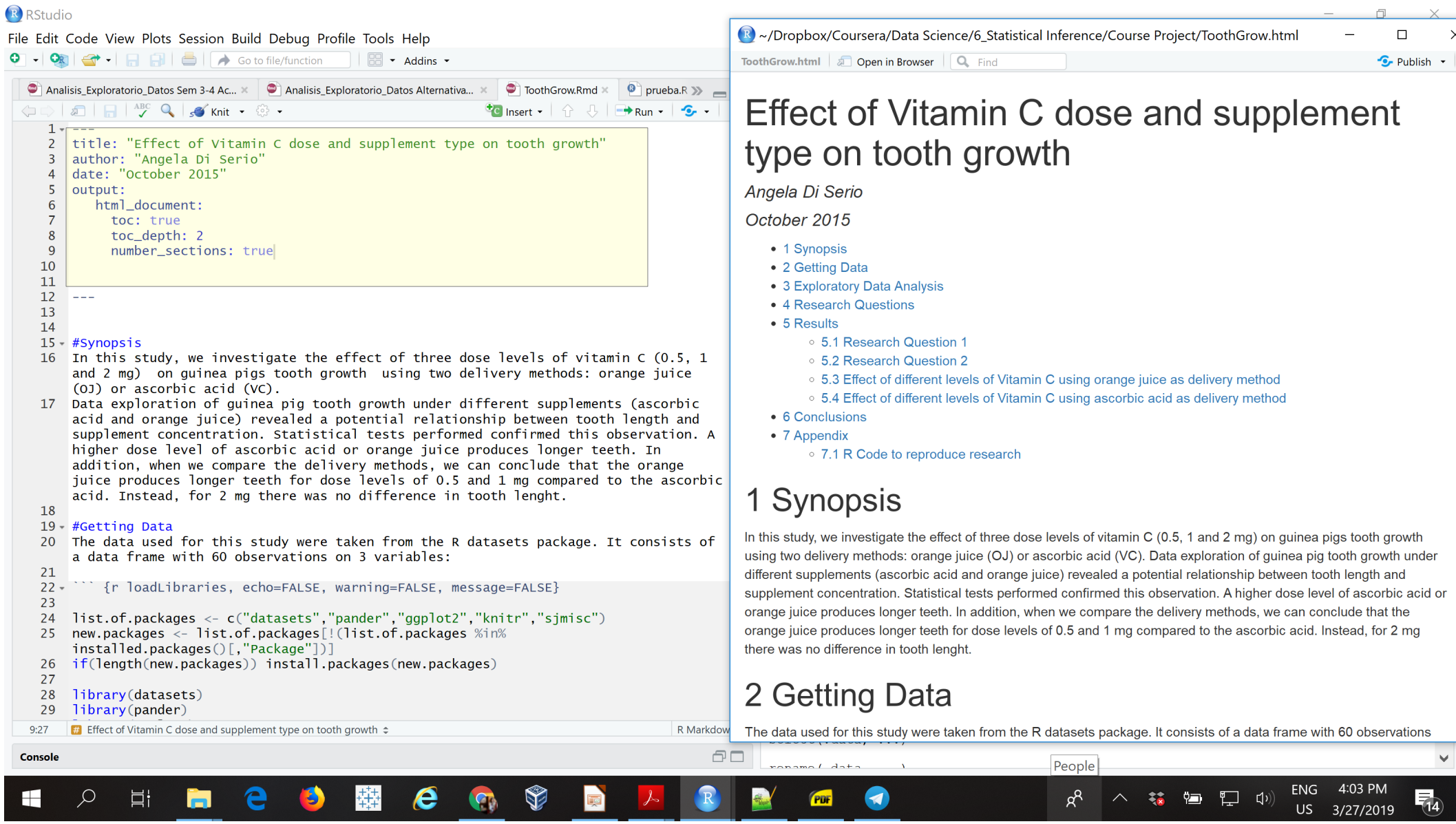
t-tests were used to compare three dose levels (0.5, 1 and 2 mg) when orange juice is used as supplement:

1. $H_o : \mu_{0.5} == \mu_1$, $H_a : \mu_{0.5} <> \mu_1$, Orange Juice
2. $H_o : \mu_{0.5} == \mu_2$, $H_a : \mu_{0.5} <> \mu_2$, Orange Juice
3. $H_o : \mu_1 == \mu_2$, $H_a : \mu_1 <> \mu_2$, Orange Juice

Since we performed multiple tests of significance, the Bonferroni correction procedure was applied to prevent a chance of making a type I error. Homogeneity of variances was checked using Levene's test confirming that

ENG US 10:52 AM 3/29/2019

6. Alternativa


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Analisis_Exploratorio_Datos Sem 3-4 Ac...
Analisis_Exploratorio_Datos Alternativa...
ToothGrow.Rmd
prueba.R
Knit
Insert
Run
---
title: "Effect of Vitamin C dose and supplement type on tooth growth"
author: "Angela Di Serio"
date: "October 2015"
output:
html_document:
toc: true
toc_depth: 2
number_sections: true

#Synopsis
In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC).
Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.
#Getting Data
The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:
```{r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}
list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
library(datasets)
library(pander)
9:27
Effect of Vitamin C dose and supplement type on tooth growth
R Markdown
Console
~/Dropbox/Coursera/Data Science/6_Statistical Inference/Course Project/ToothGrow.html
ToothGrow.html
Open in Browser
Find
Publish
Effect of Vitamin C dose and supplement type on tooth growth
Angela Di Serio
October 2015
1 Synopsis
2 Getting Data
3 Exploratory Data Analysis
4 Research Questions
5 Results
5.1 Research Question 1
5.2 Research Question 2
5.3 Effect of different levels of Vitamin C using orange juice as delivery method
5.4 Effect of different levels of Vitamin C using ascorbic acid as delivery method
6 Conclusions
7 Appendix
7.1 R Code to reproduce research
1 Synopsis
In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.
2 Getting Data
The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations
People
ENG
US
4:03 PM
3/27/2019
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

3-4 Ac... | Analisis_Exploratorio_Datos Alternativa... | ToothGrow.Rmd | prueba.R | prueba_notebook.Rmd

Knit | Insert

```
---
title: "Effect of Vitamin C dose and supplement type on tooth growth"
author: "Angela Di Serio"
date: "October 2015"
output:
   pdf_document:
     toc: true
     toc_depth: 2
     number_sections: true
   fontsize: 11pt
   geometry: margin=1in

---


#Synopsis
In this study, we investigate the effect of three dose levels of vitamin C (0.5,
mg)  on guinea pigs tooth growth  using two delivery methods: orange juice (OJ)
ascorbic acid (VC).
Data exploration of guinea pig tooth growth under different supplements (ascorbi
orange juice) revealed a potential relationship between tooth length and supplem
concentration. Statistical tests performed confirmed this observation. A higher
of ascorbic acid or orange juice produces longer teeth. In addition, when we com
delivery methods, we can conclude that the orange juice produces longer teeth fo
levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there wa
difference in tooth lenght.

#Getting Data
The data used for this study were taken from the R datasets package. It consists
frame with 60 observations on 3 variables:

```{r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}

list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
new.packages <- list.of.packages[!(list.of.packages %in%
installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)

```

14:1 (Top Level)

Console

ToothGrow.pdf

Page: 1 / 8 Find:

Effect of Vitamin C dose and supplement type on tooth growth

Angela Di Serio

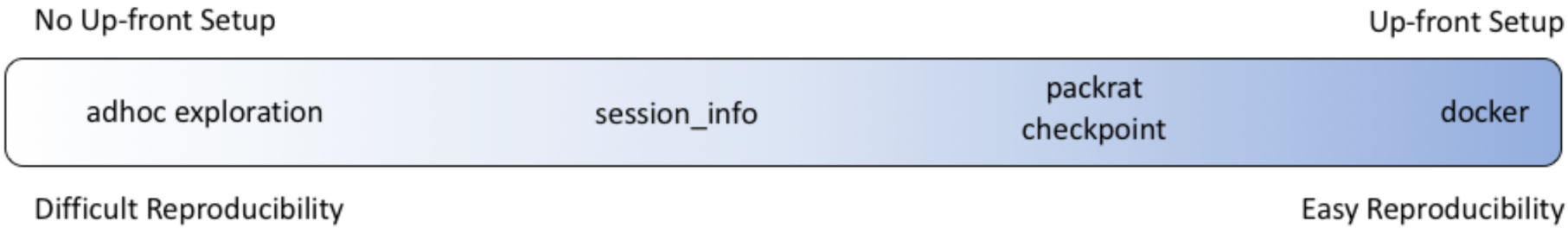
October 2015

Contents

1 Synopsis

In this study, we investigate the effect of three dose levels of vitamin C (0.5, 1 and 2 mg) on guinea pigs tooth growth using two delivery methods: orange juice (OJ) or ascorbic acid (VC). Data exploration of guinea pig tooth growth under different supplements (ascorbic acid and orange juice) revealed a potential relationship between tooth length and supplement concentration. Statistical tests performed confirmed this observation. A higher dose level of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the delivery methods, we can conclude that the orange juice produces longer teeth for dose levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no difference in tooth lenght.

2 Getting Data

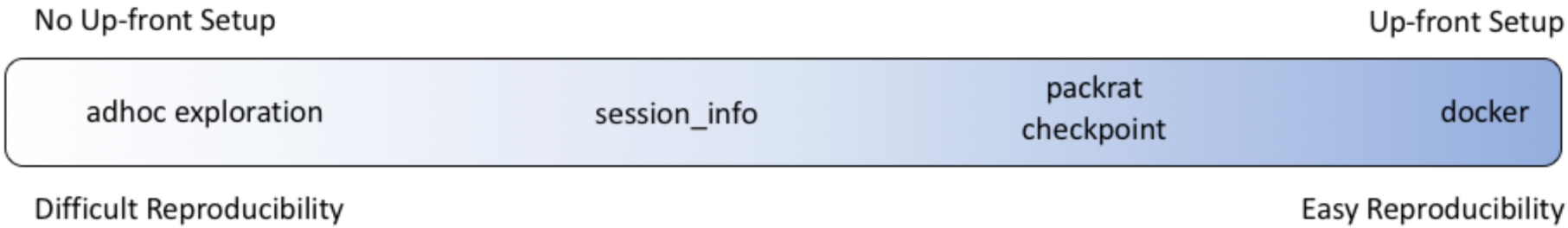
ENG US 4:07 PM 3/27/2019
```

POR QUÉ ES ÚTIL PARA IR?

Investigador inicia su proceso de programación escribiendo la prosa y el código en un simple documento R Markdown

Valores dentro de la prosa, gráficos, imágenes pueden ser generados dinámicamente basado solamente en los datos de entrada y el ecosistema de ejecución (paquetes)

POR QUÉ ES ÚTIL PARA IR?

Si otro investigador desea reproducir esta investigación o desea usar otro conjunto de datos, simplemente tiene que re-ejecutar todos los trozos de códigos o presionar el botón de *Knit* para generar el documento

El reporte es reflejo fiel de los datos de entrada utilizados y no hay intervención humana haciendo copy & paste

PROBLEMAS

La comunidad de R es muy activa

Nuevos paquetes

Actualizaciones constantes de los paquetes

GESTIÓN DE DEPENDENCIAS

Los programadores de R están familiarizados con el dolor que pueden surgir con los muchos paquetes que cambian o se vuelven obsoletos con el tiempo

La buena gestión de la dependencia garantiza que un proyecto se pueda re-calcular en otro momento o en otro ordenador

GESTIÓN DE DEPENDENCIAS

La forma en que se realice el seguimiento y manejo de las dependencias establecerá qué tan reproducible es nuestro trabajo



Cole Arendt. Package Management for Reproducible R Code
https://rviews.rstudio.com/2018/01/18/package-management-for-reproducible-r-code/

Relevance

```
rel.tradicional<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Traditional Method"]
rel.tradicional
rel.AR<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Augmented Reality"]
rel.AR
var.test(rel.tradicional, rel.AR)
t.test(rel.tradicional, rel.AR, var.equal = F)
t.test(rel.tradicional, rel.AR, var.equal = T)
```

```
> rel.tradicional<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Traditional Method"]
> rel.tradicional
 [1] 2.88 3.33 3.33 2.67 3.75 3.22 3.33 3.63 3.22 3.00 2.78 3.44 4.13 2.89 3.00 2.22 3.75 4.00 3.11 3.44 3.56 2.78 2.67 3.33 3.56 3.78 4.22 3.00 3.50
[30] 3.22 4.33 3.25 2.33 3.11 3.44 3.33 3.78 2.44 3.00 3.89 3.22 3.78 2.89 3.44 2.67 2.33 3.44 3.00 3.56 3.44 2.67 3.33 3.89 3.22 4.00 3.44 3.56 2.89
[59] 3.56
> rel.AR<-sinAtipicos$Relevance[sinAtipicos$cuestionario=="Augmented Reality"]
> rel.AR
 [1] 3.89 3.11 3.56 3.33 3.56 3.33 2.78 3.67 3.11 3.33 3.33 3.00 3.44 4.00 3.78 4.22 2.67 3.89 3.33 3.78 3.71 3.00 3.33 4.56 4.22 4.00 3.11 3.56 3.22
[30] 4.25 2.38 3.78 3.44 4.22 3.67 3.56 3.33 3.22 4.44 3.56 3.67 3.11 3.50 4.67 3.56 2.78 4.00 3.78 3.11 2.56 2.63 3.67 3.33 2.67 3.33 4.33 3.22 4.33
[59] 4.44 3.44 3.13 4.33

> var.test(rel.tradicional, rel.AR)

        F test to compare two variances

data:  rel.tradicional and rel.AR
F = 0.8072, num df = 58, denom df = 61, p-value = 0.413
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4839619 1.3508779
sample estimates:
ratio of variances
         0.8072067
```

p-grande no se puede rechazar la hipotesis nula y se concluye que las varianzas son iguales
Aplicamos t-test asumiendo que las varianzas son iguales

```
> t.test(rel.tradicional, rel.AR, var.equal = T)

        Two Sample t-test

data:  rel.tradicional and rel.AR
t = -2.7128, df = 119, p-value = 0.007662
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.43044524 -0.06720375
sample estimates:
mean of x mean of y
 3.287627  3.536452
```


GESTIÓN DE DEPENDENCIAS

La forma en que se realice el seguimiento y manejo de las dependencias establecerá qué tan reproducible es nuestro trabajo



Cole Arendt. Package Management for Reproducible R Code
https://rviews.rstudio.com/2018/01/18/package-management-for-reproducible-r-code/

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

atorio_Datos Alternativa... | prueba.R | prueba_notebook.Rmd | ToothGrow.Rmd | pruebaPython.Rmd | mydfs

Knit | Insert | Run

```
concentration. Statistical tests performed confirmed this observation. A higher dose level
of ascorbic acid or orange juice produces longer teeth. In addition, when we compare the
delivery methods, we can conclude that the orange juice produces longer teeth for dose
levels of 0.5 and 1 mg compared to the ascorbic acid. Instead, for 2 mg there was no
difference in tooth lenght.

\newpage


#Session Information

```{r loadLibraries, echo=FALSE, warning=FALSE, message=FALSE}

list.of.packages <- c("datasets","pander","ggplot2","knitr","sjmisc")
new.packages <- list.of.packages[!(list.of.packages %in%
installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)

library(datasets)
library(pander)
library(ggplot2)
library(knitr)
library(sjmisc)

```

```{r}
sessionInfo()
```


#Getting Data
The data used for this study were taken from the R datasets package. It consists of a data
frame with 60 observations on 3 variables:
```

20:1 | Synopsis | R Markdow

Console

ToothGrow.pdf

Page: 1 / 8 Find:

Session Information

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17134)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] sjmisc_2.7.9  knitr_1.22    ggplot2_3.1.0 pander_0.6.3
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.0        pillar_1.3.1      plyr_1.8.4
##  [4] forcats_0.4.0     tools_3.3.2       digest_0.6.12
##  [7] evaluate_0.13     tibble_2.0.1      gtable_0.2.0
## [10] pkgconfig_2.0.2   rlang_0.3.1       yaml_2.2.0
## [13] haven_2.1.0       xfun_0.5          withr_2.1.2
## [16] dplyr_0.8.0.1     stringr_1.2.0     hms_0.4.2
## [19] sjlabelled_1.0.17 grid_3.3.2        tidyselect_0.2.5
## [22] glue_1.3.0        R6_2.2.2          rmarkdown_1.12
## [25] purrr_0.2.5       magrittr_1.5      scales_1.0.0
## [28] htmltools_0.3.6   assertthat_0.2.0  insight_0.1.2
## [31] colorspace_1.3-2  stringi_1.1.6     lazyeval_0.2.1
## [34] munsell_0.5.0     crayon_1.3.4
```

Getting Data

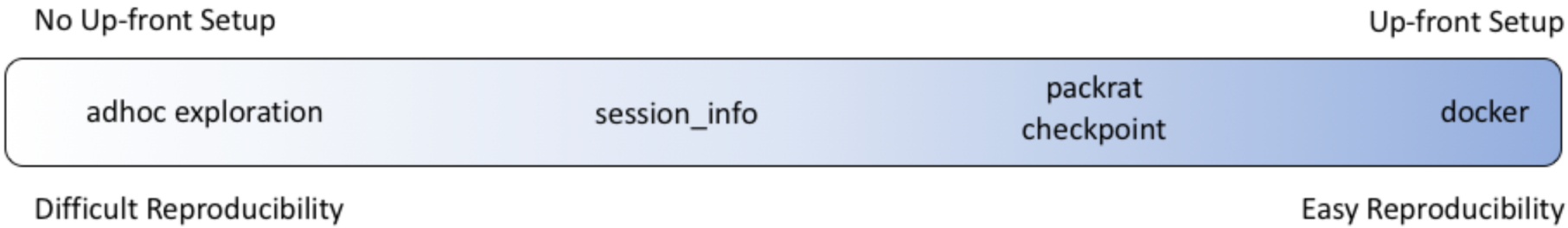
The data used for this study were taken from the R datasets package. It consists of a data frame with 60 observations on 3 variables:

Table 1. ToothGrowth Dataset variables

ENG US 4:01 PM 3/28/2019

GESTIÓN DE DEPENDENCIAS

La forma en que se realice el seguimiento y manejo de las dependencias establecerá qué tan reproducible es nuestro trabajo



Cole Arendt. Package Management for Reproducible R Code
https://rviews.rstudio.com/2018/01/18/package-management-for-reproducible-r-code/

PACKRAT
Gestor de Dependencias

AISLAR

1. La instalación de nuevos paquetes o la actualización de paquetes para un proyecto no interferirá con otros proyectos en desarrollo y viceversa

2. Packrat le da a cada proyecto su propia biblioteca privada de paquetes

PACKRAT
Gestor de Dependencias

PORTABILIDAD

1. El proyecto puede ser transportado de un ordenador a otro, incluso a diferentes plataformas

2. Packrat facilita la instalación de los paquetes de los que depende el proyecto

PACKRAT
Gestor de Dependencias

REPRODUCIBILIDAD

1 Packrat registra las versiones exactas de los paquetes de las que depende el proyecto, y garantiza que esas versiones exactas sean las que se instalen

PACKRAT
Gestor de Dependencias

FLUJO DE TRABAJO

1. packrat :: init( ) para iniciar packrat mientras se desarrolla el análisis
2. packrat :: snapshot( ) para almacenar la lista de paquetes en uso, sus fuentes y versiones
3. packrat :: restore( ) aplica la “instantánea” más reciente a la biblioteca privada del proyecto
4. packrat :: bundle( ) comprime un proyecto packrat para compartir fácilmente
5. packrat :: unbundle( ) descomprime un proyecto packrat

6. Alternativa

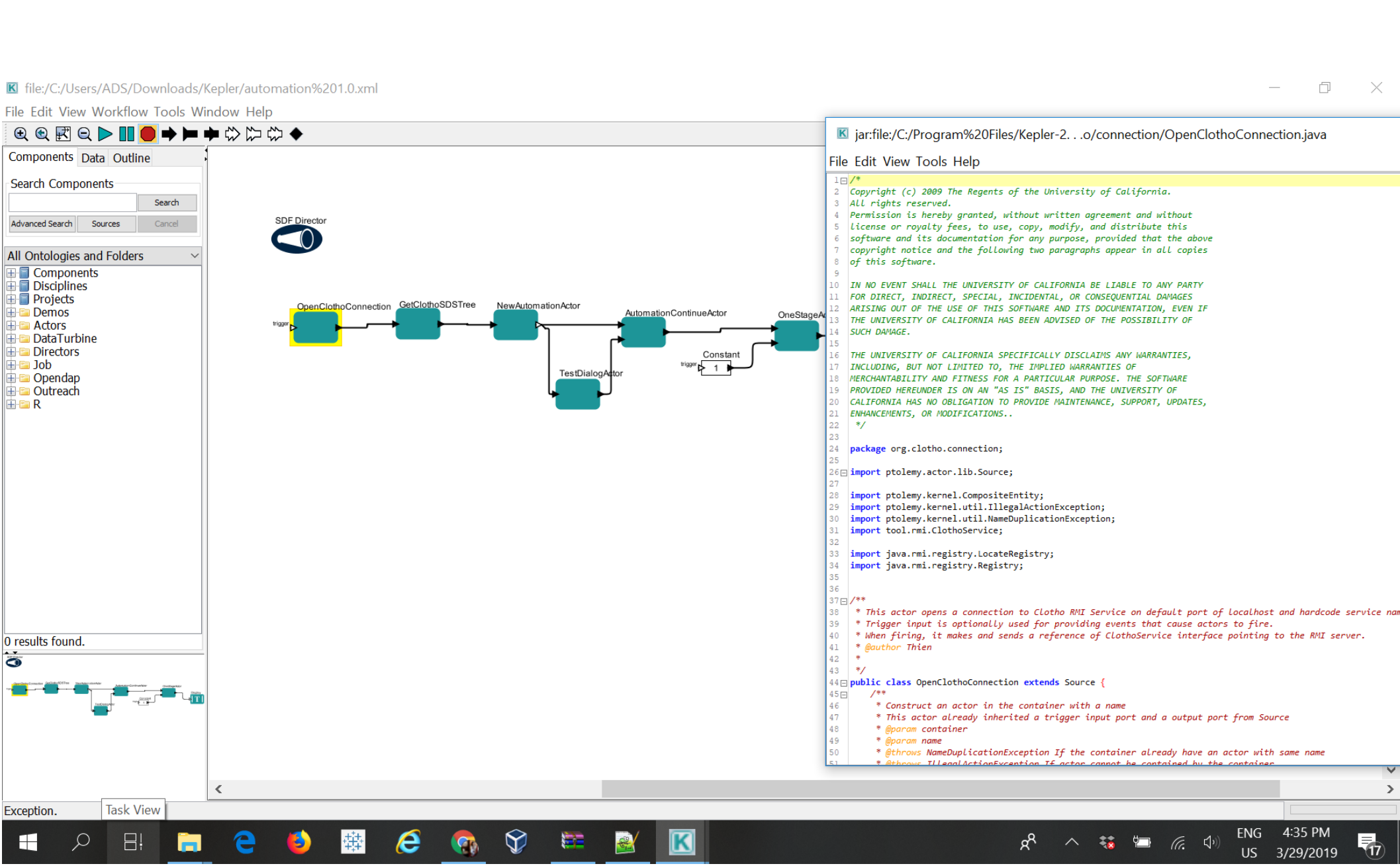
OTRAS ALTERNATIVAS

Notebooks: IPython, R Notebooks

Workflow Management:

- Proyecto Kepler (Java, R)
- Doit (python)
- Snakemake (python)

makefile



6. Alternativa

PARA CONCLUIR

CHECKLIST

Pensar en el pipeline completo: es posible reproducir todas las etapas?

El proceso de limpieza y análisis de los datos está automatizado?

- Estoy realizando cosas “a mano”, editar tablas, figuras, formateando datos?
- Lo que estoy haciendo soporta scripts? En caso contrario puedo diseñar una descripción detallada del proceso?

CHECKLIST

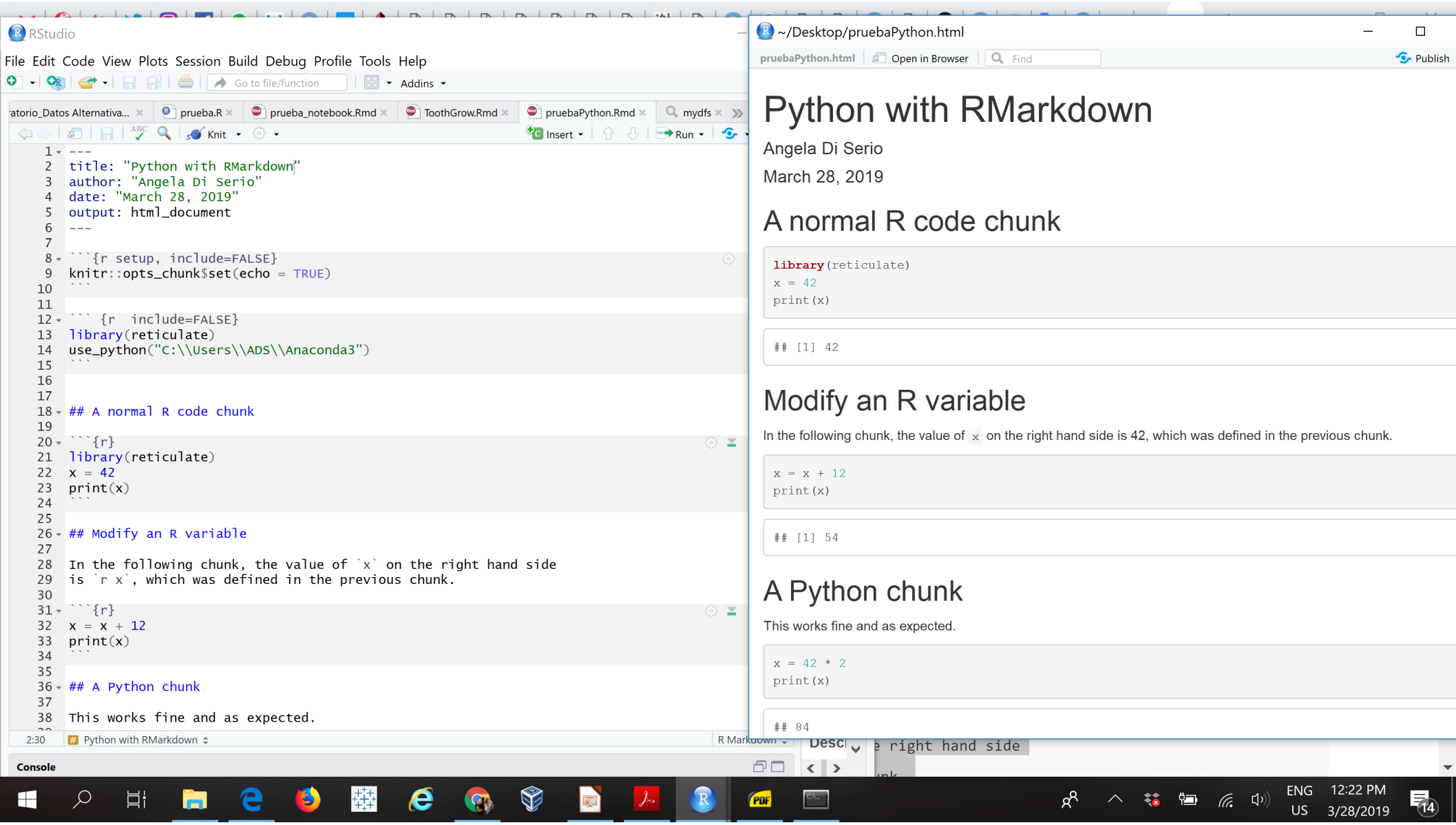
Estoy usando control de versiones?

Estoy documentando lo relacionado con el software y hardware utilizado?

- Arquitectura
- Sistema de operación
- Librerías usadas, paquetes, versiones de las mismas

Estoy salvando los archivos de datos necesarios?

GRACIAS



3. Investigación Reproducible