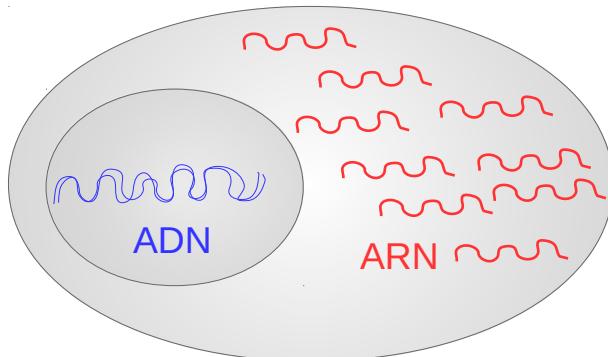
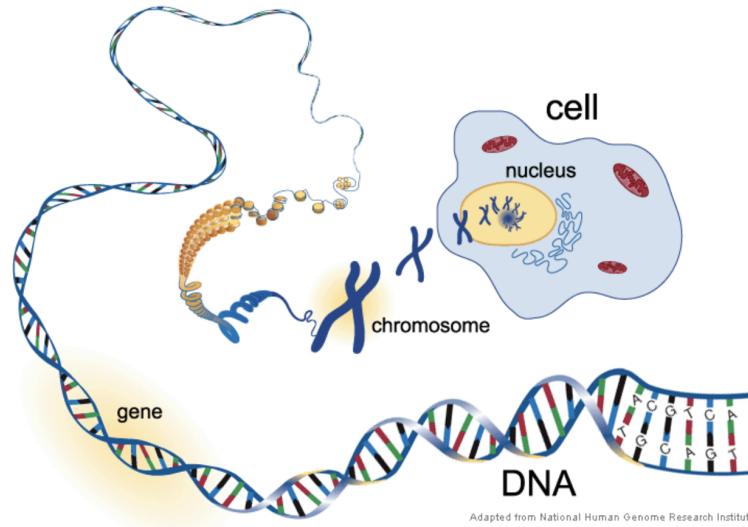


Código y paquetes para el análisis y visualización de secuencias biológicas en R

 @fdiazviraque



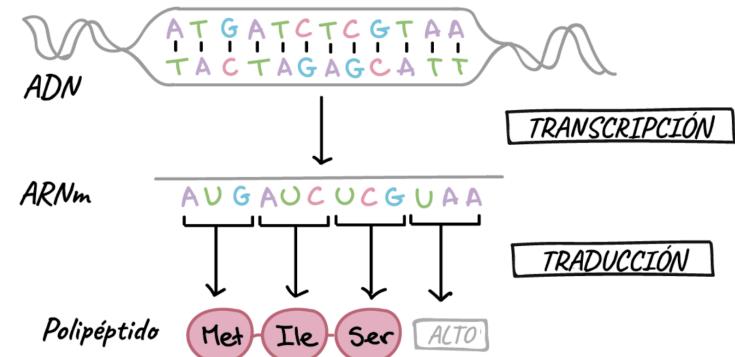
> head (data)



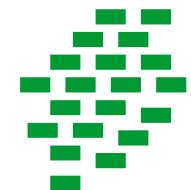
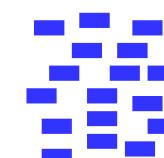
[[1]][1]

[[1]][2]

EL DOGMA CENTRAL



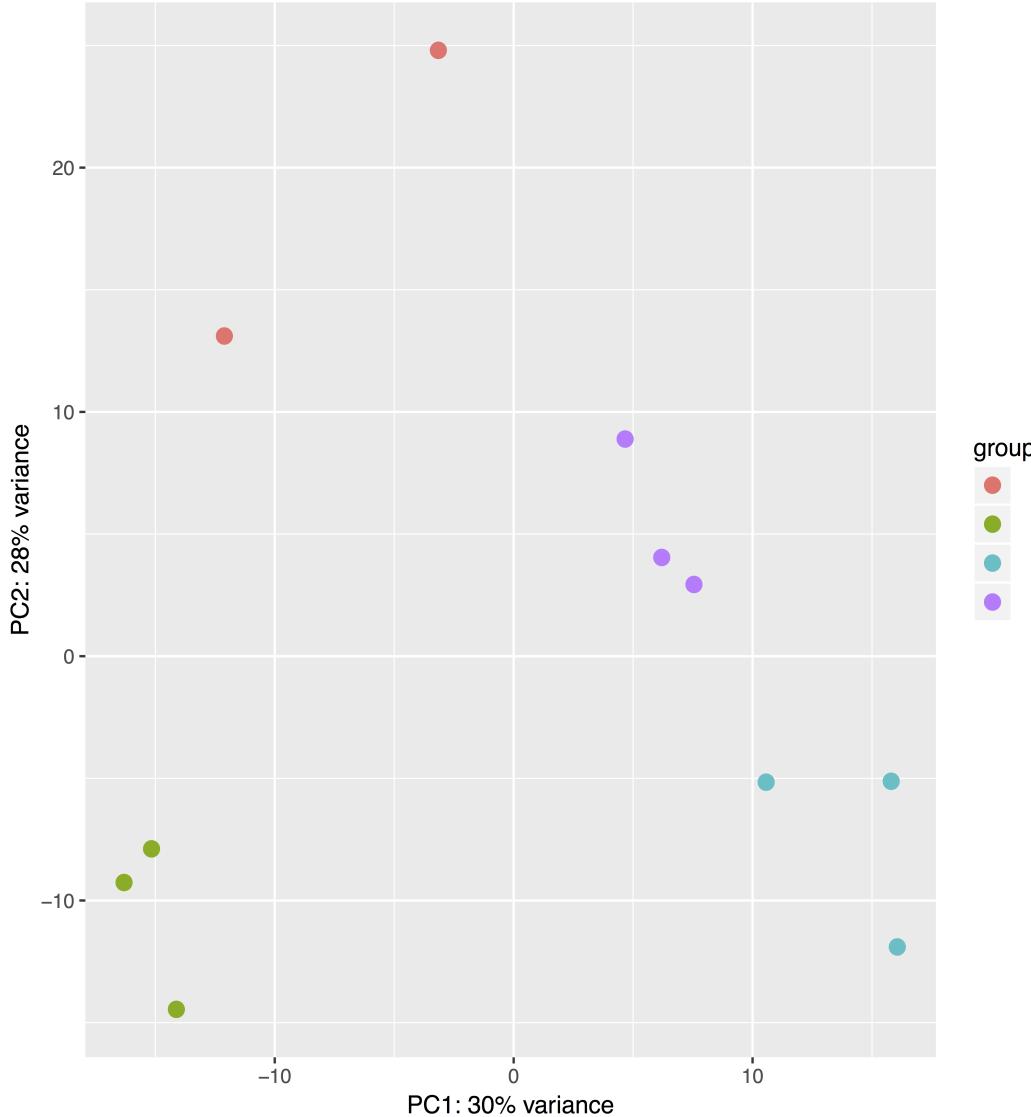
ACGTACACACGTACACACCGCTGTAGCATAACACGTTACACACAAGTAGTAACAC



Composición y abundancia relativa

Variabilidad de los datos

Análisis de componentes principales



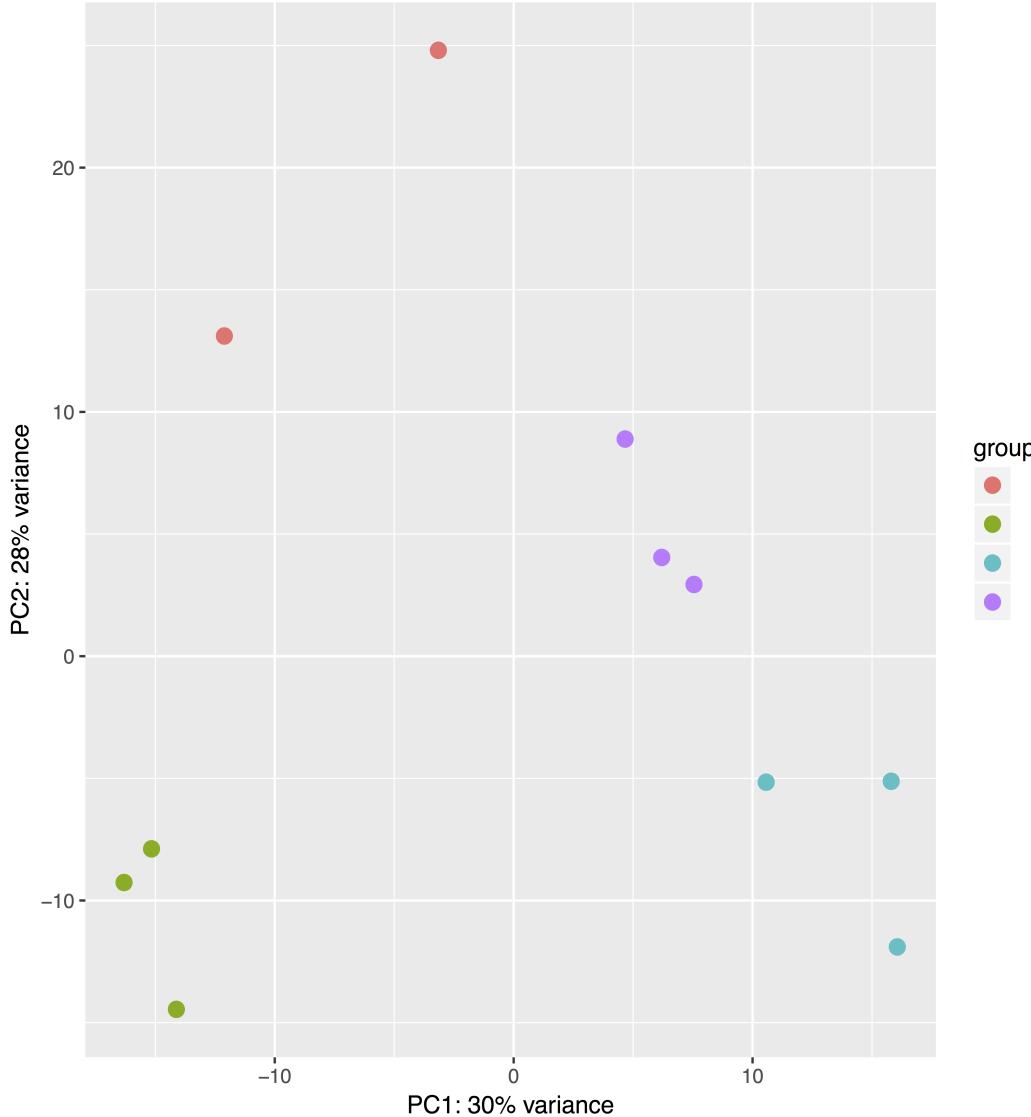
DESeq2::plotPCA

PCA functions in R :

prcomp() (**stats**)
princomp() (**stats**)
PCA() (**FactoMineR**)
dudi.pca() (**ade4**)
acp() (**amap**)

Variabilidad de los datos

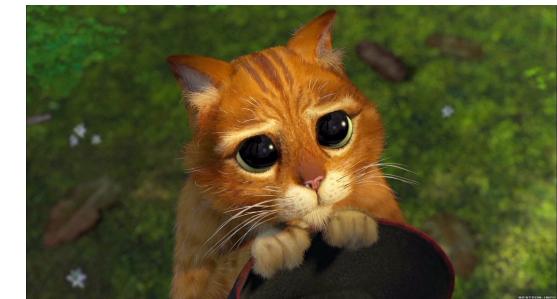
Análisis de componentes principales



DESeq2::plotPCA

PCA functions in R :

prcomp() (**stats**)
princomp() (**stats**)
PCA() (**FactoMineR**)
dudi.pca() (**ade4**)
acp() (**amap**)



Perdón por las gráficas a ciegas

Variabilidad de los datos

Estabilización de la varianza en los datos

Muchos métodos estadísticos para análisis exploratorios de datos multidimensionales funcionan mejor para datos que tienen el mismo rango de varianza en diferentes valores de la media.

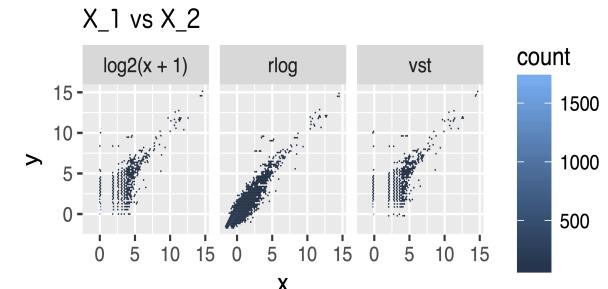
En datos de RNA la varianza esperada crece con la media.

Y si uno realiza PCA directamente en una matriz de conteos la gráfica depende principalmente de los genes con conteos más altos porque muestran las mayores diferencias absolutas entre las muestras.

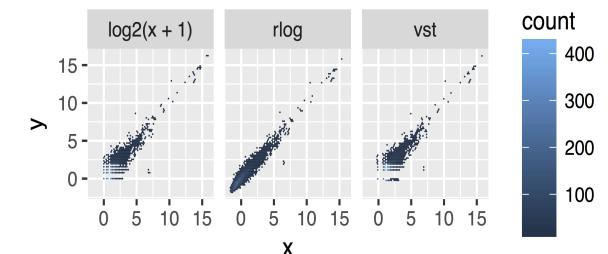
Generalmente $\log_2(x+1) \rightarrow$ los genes con los conteos más bajos contribuirán con mucho ruido, porque tomar el logaritmo de conteos pequeños en realidad “infla” su varianza.

Los genes de conteos bajos (con baja relación señal / ruido) contribuirán demasiado a las distancias muestra-muestra y PCA.

Scatterplot of transformed data



Z_2 vs Z_3



Variabilidad de los datos

```
# Estabilización de la varianza en los datos
```

```
# DESeq2::rlog ('regularized log' transformation)
```

```
# DESeq2::vst (varianceStabilizingTransformation)
```

Darán un resultado similar a la transformación log2 para los conteos altos mientras que para los genes con conteos bajos los valores se reducen a los promedios de los genes en todas las muestras.

```
# Effect of the transformation plot
```

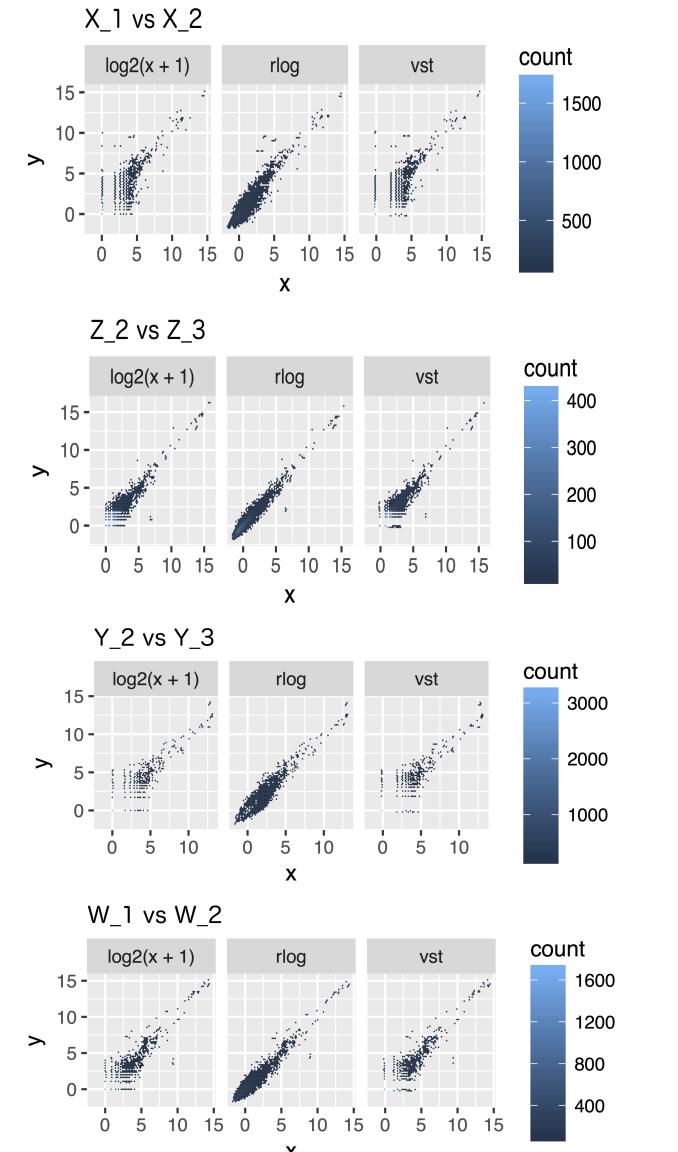
```
library("dplyr")
library("ggplot2")
```

```
> df <- bind_rows(
  as_data_frame(data %>% mutate(transformation = "log2(x + 1)"),
  as_data_frame(data %>% mutate(transformation = "rlog"),
  as_data_frame(data %>% mutate(transformation = "vst")))
```

```
> colnames(df)[1:2] <- c("x", "y")
```

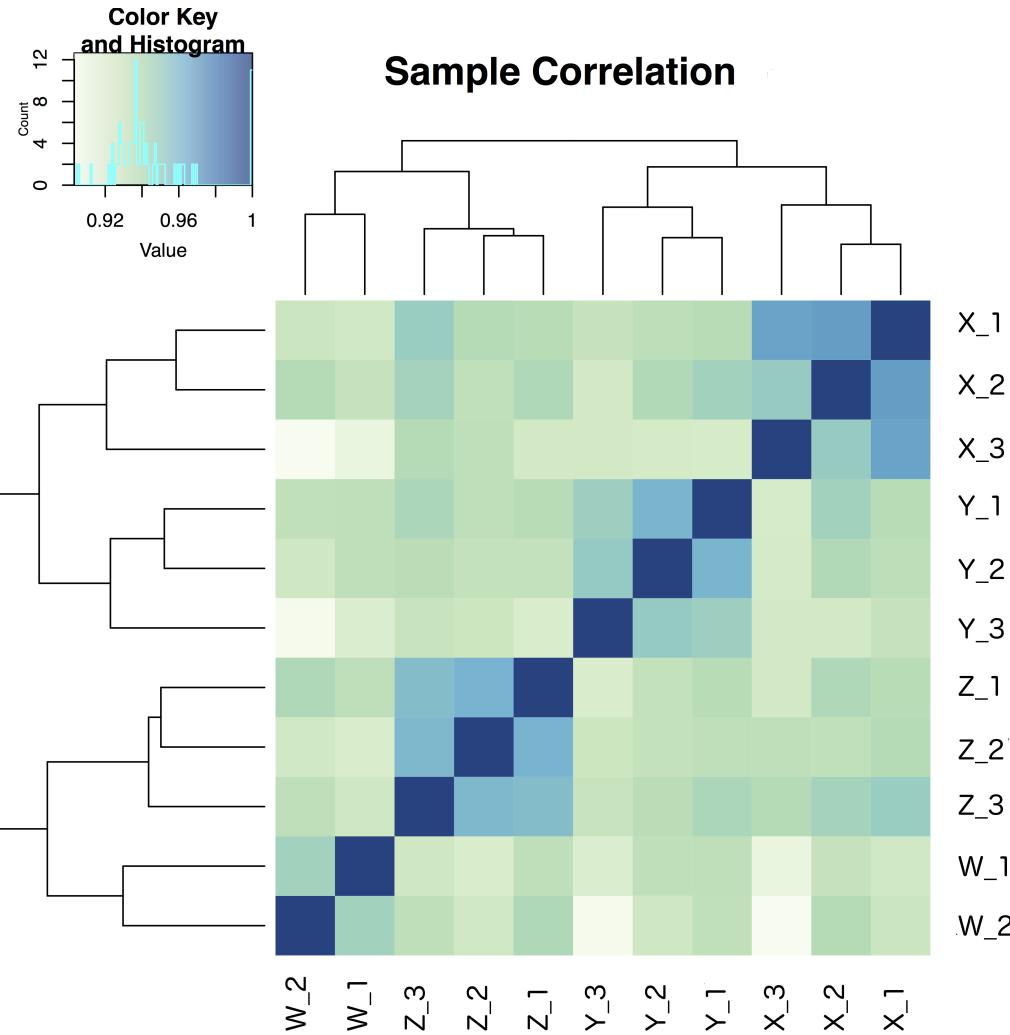
```
> ggplot(df, aes(x = x, y = y)) + geom_hex(bins = 80) +
  coord_fixed() + facet_grid(. ~ transformation) +
  theme(plot.title = element_text(size=12))
```

```
# Scatterplot of transformed data
```



Variabilidad de los datos

Correlación, varianza & covarianza



```
# cor()
# var()
# cov()
```

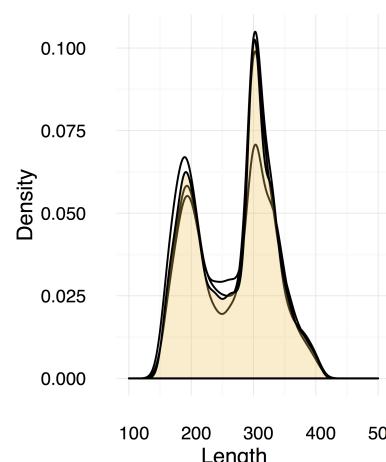
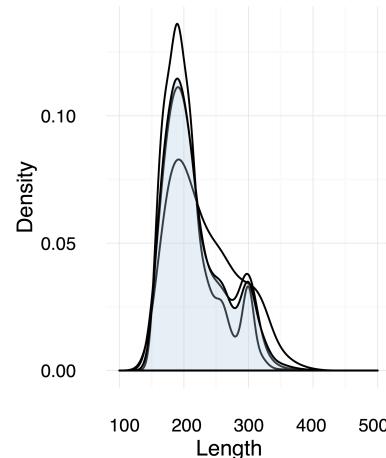
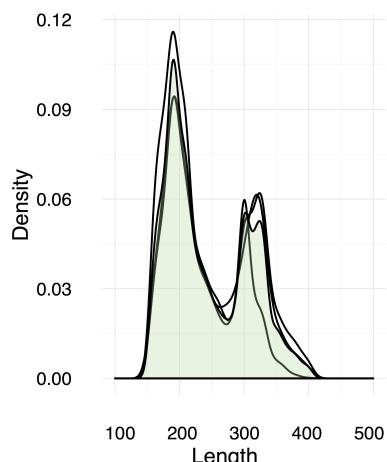
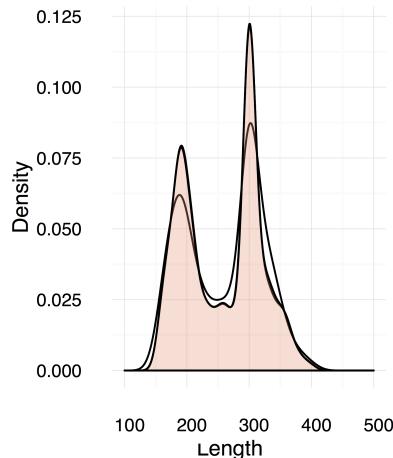
```
# gplots::heatmap.2
```

```
> heatmap.2(cor(data), trace="none",
  col=hmc.col, main="Sample Correlation")
```

Un **heatmap** es una representación gráfica de los datos que usa un sistema de codificación en colores para representar diferentes valores

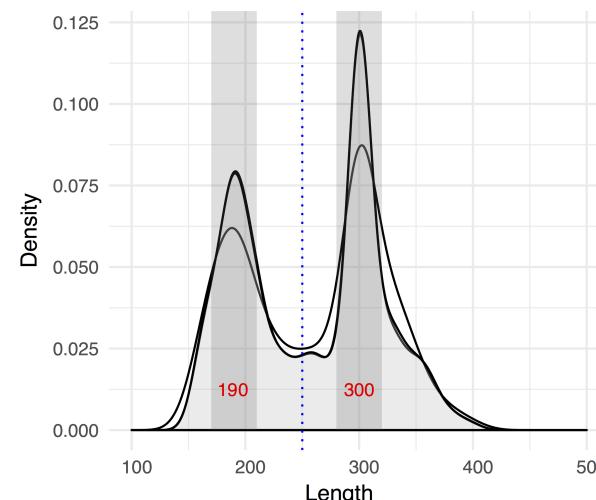


Distribución de largos de los datos



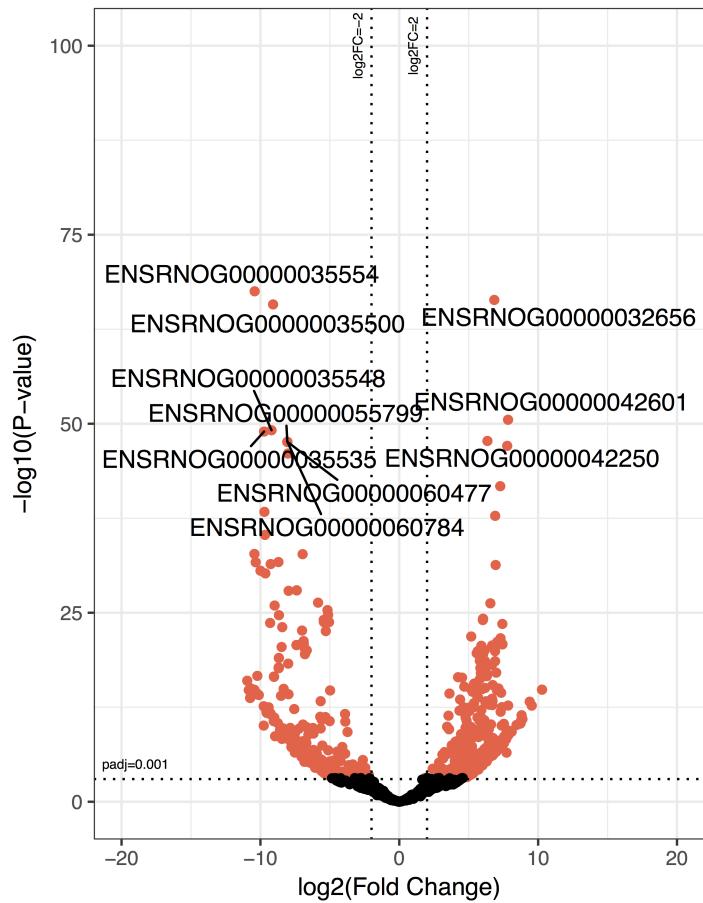
```
> a <- ggplot() +  
  geom_density(data=df, aes(x=x1)) +  
  geom_density(data=df, aes(x=x2)) +  
  geom_density(data=df2, aes(x=x3, fill=cond), alpha=.3) +  
  labs(title = "Muestra X") +  
  labs(y="Density") + labs(x="Length") + xlim(100, 500) +  
  scale_fill_manual( values = c("#fc9272")) +  
  theme(legend.position="none") +  
  theme_minimal() + labs(fill = "Sample")  
  
#geom_vline(xintercept = 250, linetype="dotted", color = "blue", size=.5) #+  
#annotate("rect", xmin = 170, xmax = 210, ymin = 0, ymax = Inf, alpha = .2) +  
#annotate("rect", xmin = 280, xmax = 320, ymin = 0, ymax = Inf, alpha = .2) +  
#annotate("text",x=190,y=0.0125, label = "190",size=3,colour = "red") +  
#annotate("text",x=300,y=0.0125, label = "300",size=3,colour = "red")
```

```
> myplot <- grid.arrange(a,b,c,d, top=textGrob("Length distribution"))  
> ggsave(filename="length_dist.pdf", plot=myplot, width=8, height=8)
```



Volcano plots

Volcano plot



```
library("ggplot2")
library("ggrepel") #Avoid overlapping labels

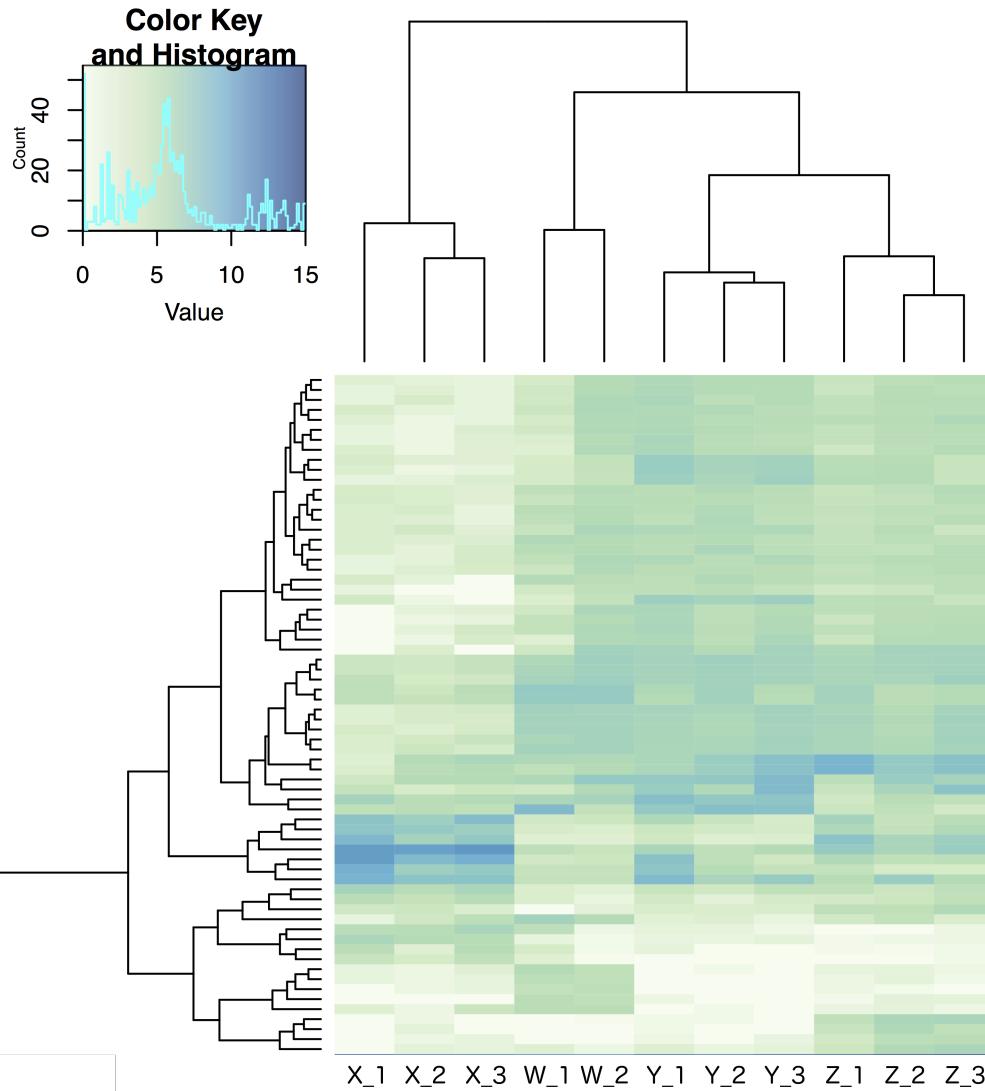
> mutatedddf <- mutate (df, sig=ifelse(res_df$padj<0.001,
  "padj<0.001", "padj>0.001")) # different colors depending on significance

> input <- cbind(gene=rownames(res_df), mutatedddf )

> volc = ggplot(input, aes(log2FoldChange, -log10(pvalue))) +
  geom_point(aes(col=sig)) +
  scale_color_manual(values=c("tomato", "black")) +
  xlim(-20,20) +
  labs(x = "log2(Fold Change)", y = "-log10(P-value)") +
  ggtitle("Volcano plot") +
  geom_vline(xintercept = 2, linetype="dotted", color = "black", size=.5) +
  geom_vline(xintercept = -2, linetype="dotted", color = "black", size=.5) +
  geom_hline(yintercept = 3, linetype="dotted", color = "black", size=.5) +
  annotate("text",x=-19,y=7, label = "padj=0.001",size=3) +
  annotate("text",x=1,y=173, label = "log2FC=2",size=2.8, angle = 90) +
  annotate("text",x=-3,y=173, label = "log2FC=-2",size=2.8, angle = 90)

> volc+geom_text_repel(data=head(input, 10), aes(label=gene)) + theme_bw()
> ggsave("Volcanoplot.pdf", device="pdf")
```

Heatmaps (Qué y Cuánto)

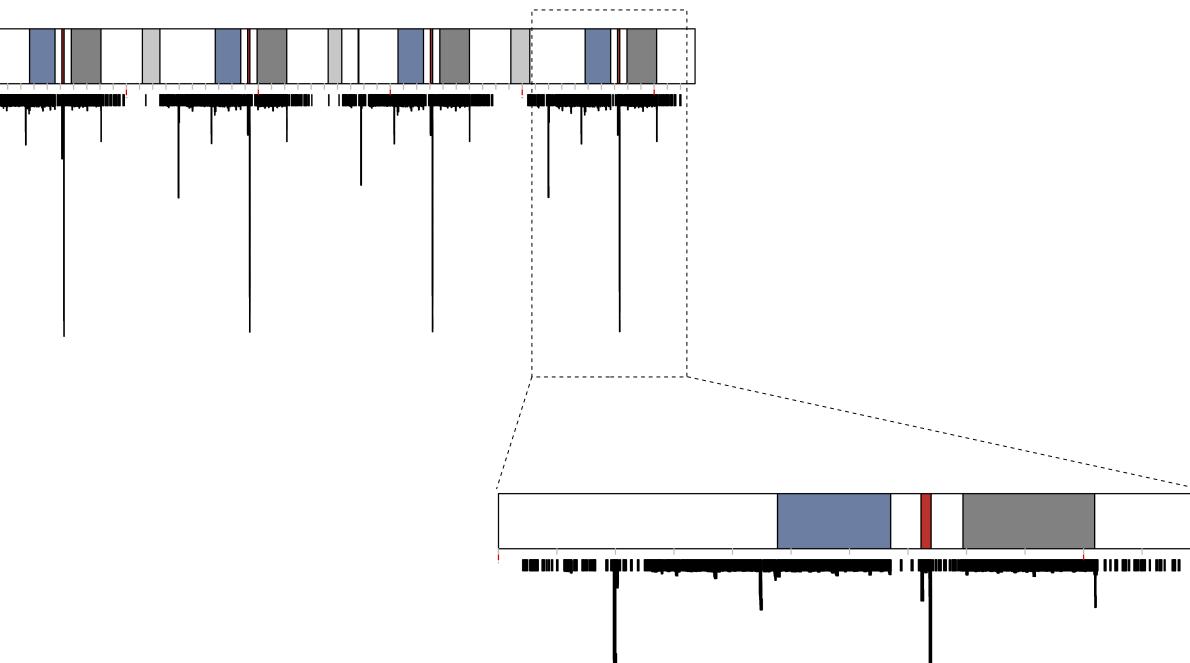


Un **heatmap** es una representación gráfica de los datos que usa un sistema de codificación en colores para representar diferentes valores.

gplots::heatmap.2

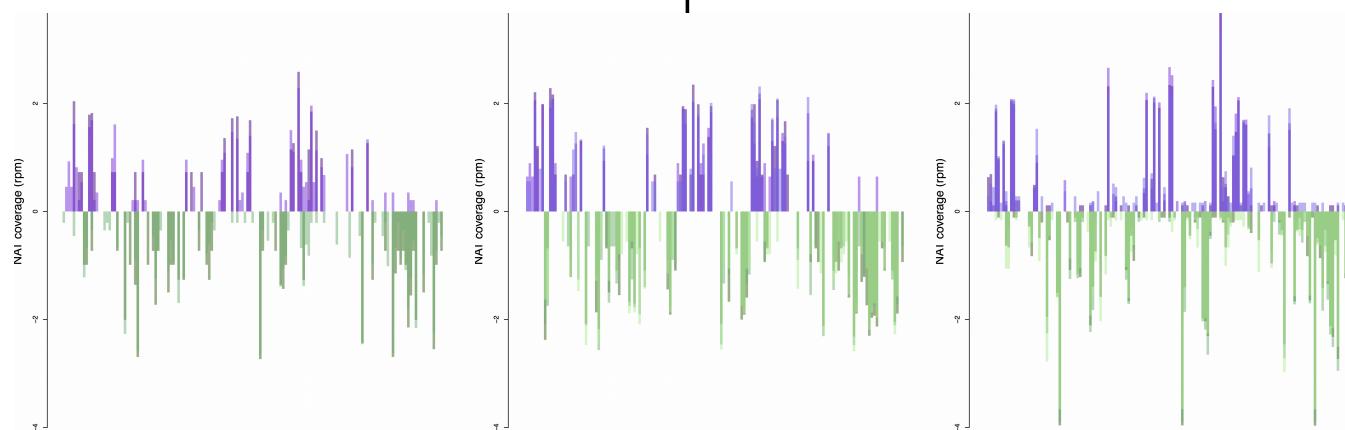
```
> hmcol = colorRampPalette(brewer.pal(9, "GnBu"))(100)
> heatmap.2(data, col=hmcol, trace="none",
  labCol=samplenames, margin=c(5,10), cexRow=0.3,
  cexCol=0.5)
```

Visualización



```
# library("karyoploteR")
# library("seqinr")
# library("GenomicRanges")
```

```
GRanges object with 3 ranges and 2 metadata columns:
 seqnames      ranges strand |   score      GC
      <Rle>    <IRanges> <Rle> | <integer> <numeric>
 txA       chr2    103-106    + |      5     0.45
 txB       chr1    107-109    + |      3     0.3
 txB       chr1    113-115   - |      4     0.5
 -----
 seqinfo: 2 sequences from an unspecified genome; no seqlengths
```

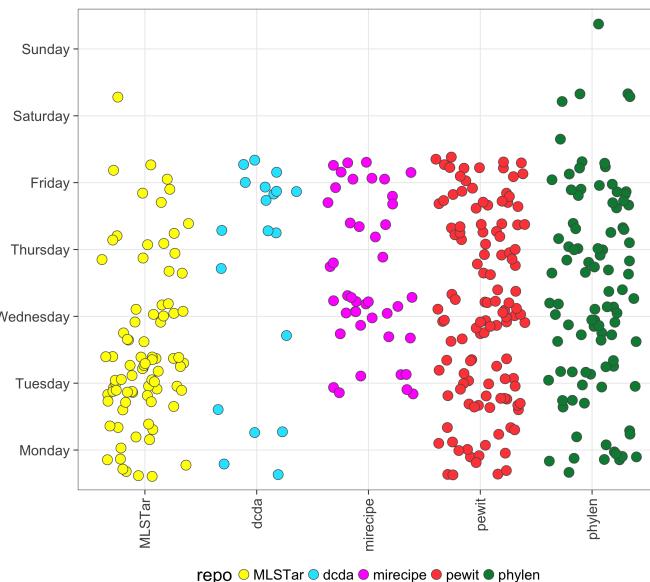




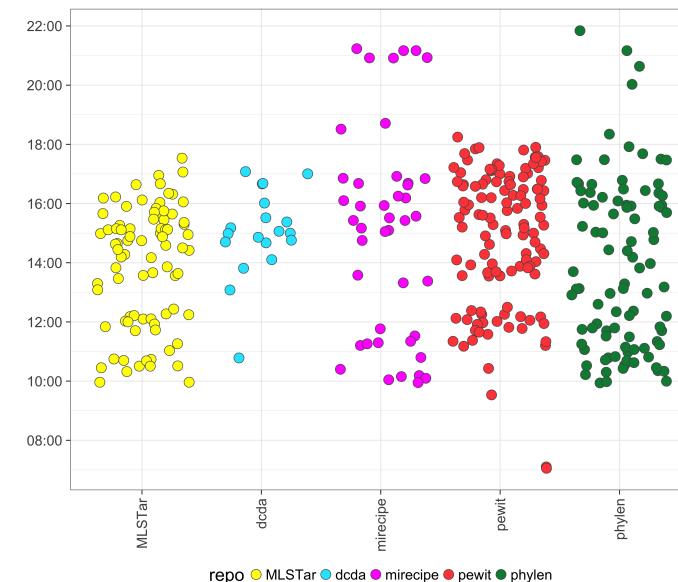
Cuantas horas trabaja Nacho?



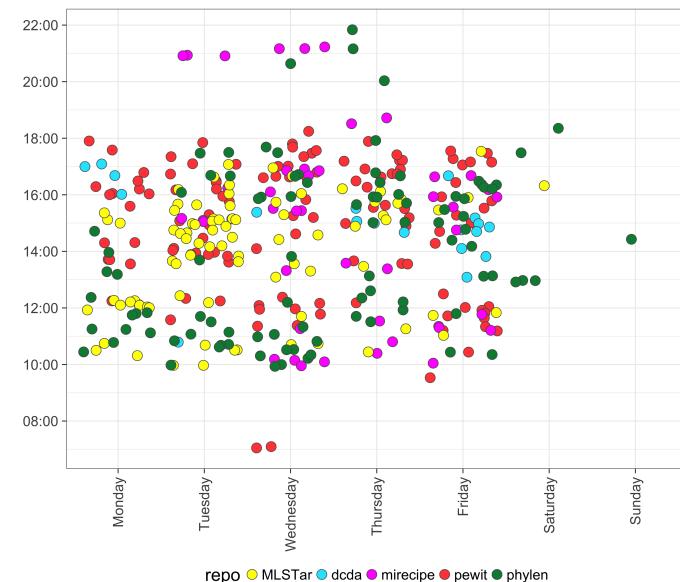
Post de Dean Attali en **DataCamp** community



Los lunes a Nacho no le gusta trabajar en mirceipe



Nacho le dedica un día menos en la semana a mirceipe pero le dedica mas horas



A Nacho le cuesta arrancar los lunes



Que tan productivo es Hadley Wickham?



Hadley Wickham
hadley

📍 Houston, TX
✉️ Sign in to view email
🔗 <http://hadley.nz>

Organizations



Overview Repositories 208 Stars 112 Followers 12.7k Following 6

Pinned repositories

[tidyverse/ggplot2](#)

An implementation of the Grammar of Graphics in R

● R ★ 3k ⚡ 1.2k

[tidyverse/dplyr](#)

Dplyr: A grammar of data manipulation

● R ★ 2.3k ⚡ 884

[tidyverse/tidyverse](#)

Easily install and load packages from the tidyverse

● R ★ 505 ⚡ 110

[r4ds](#)

R for data science

● R ★ 1.2k ⚡ 1.6k

[r-lib/devtools](#)

Tools to make an R developer's life easier

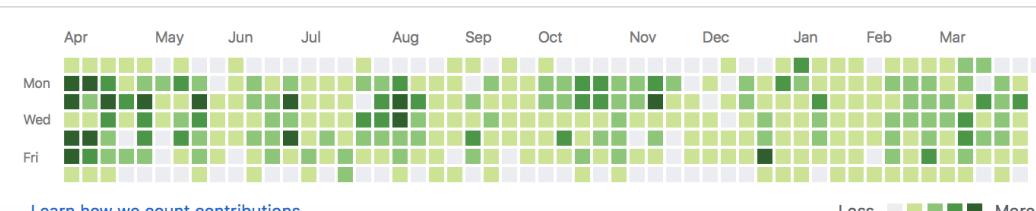
● R ★ 1.6k ⚡ 574

[adv-r](#)

Advanced R programming: a book

● TeX ★ 1.2k ⚡ 1.2k

4,190 contributions in the last year

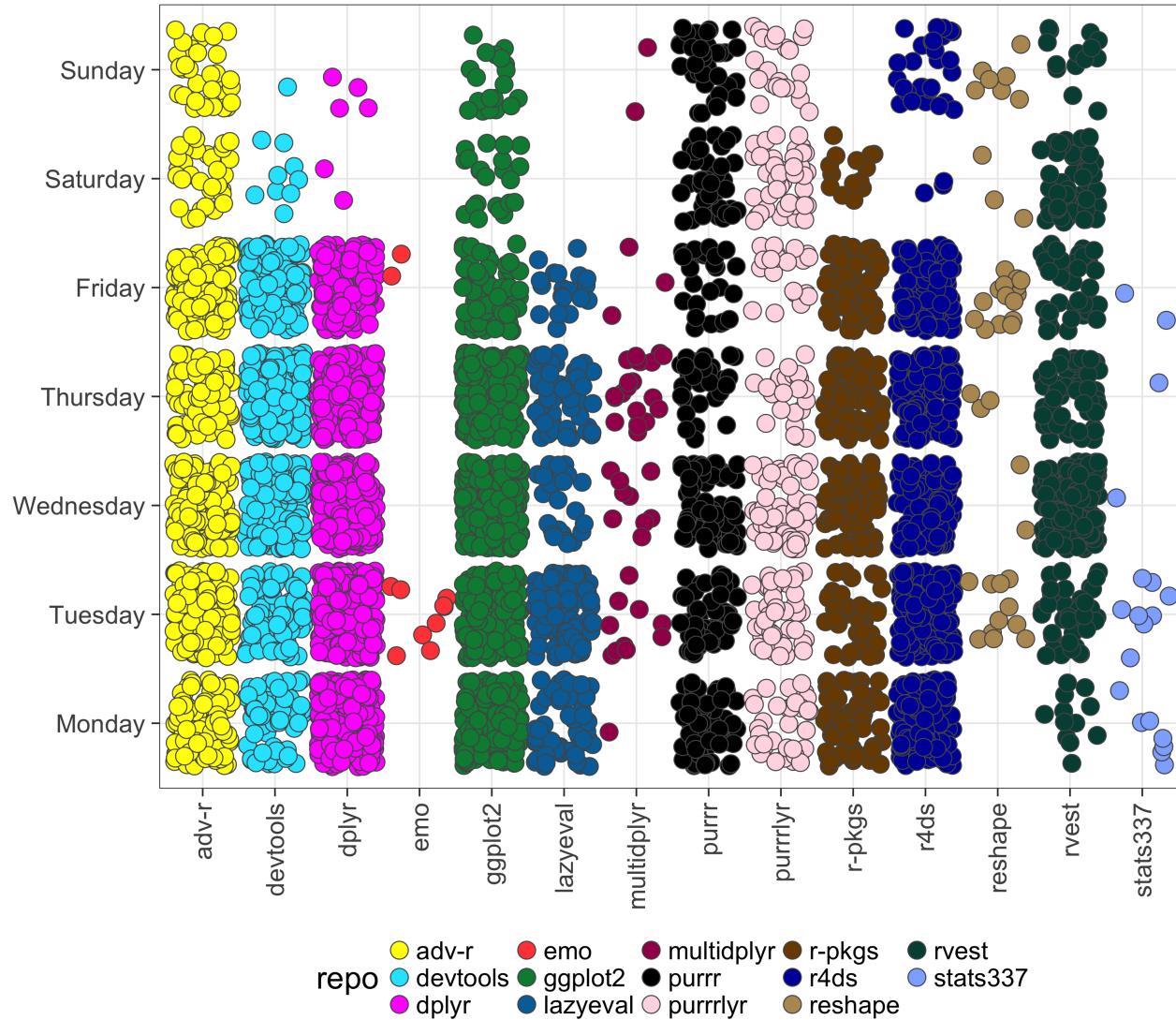


 @hadleywickham





Que tan productivo es Hadley Wickham?



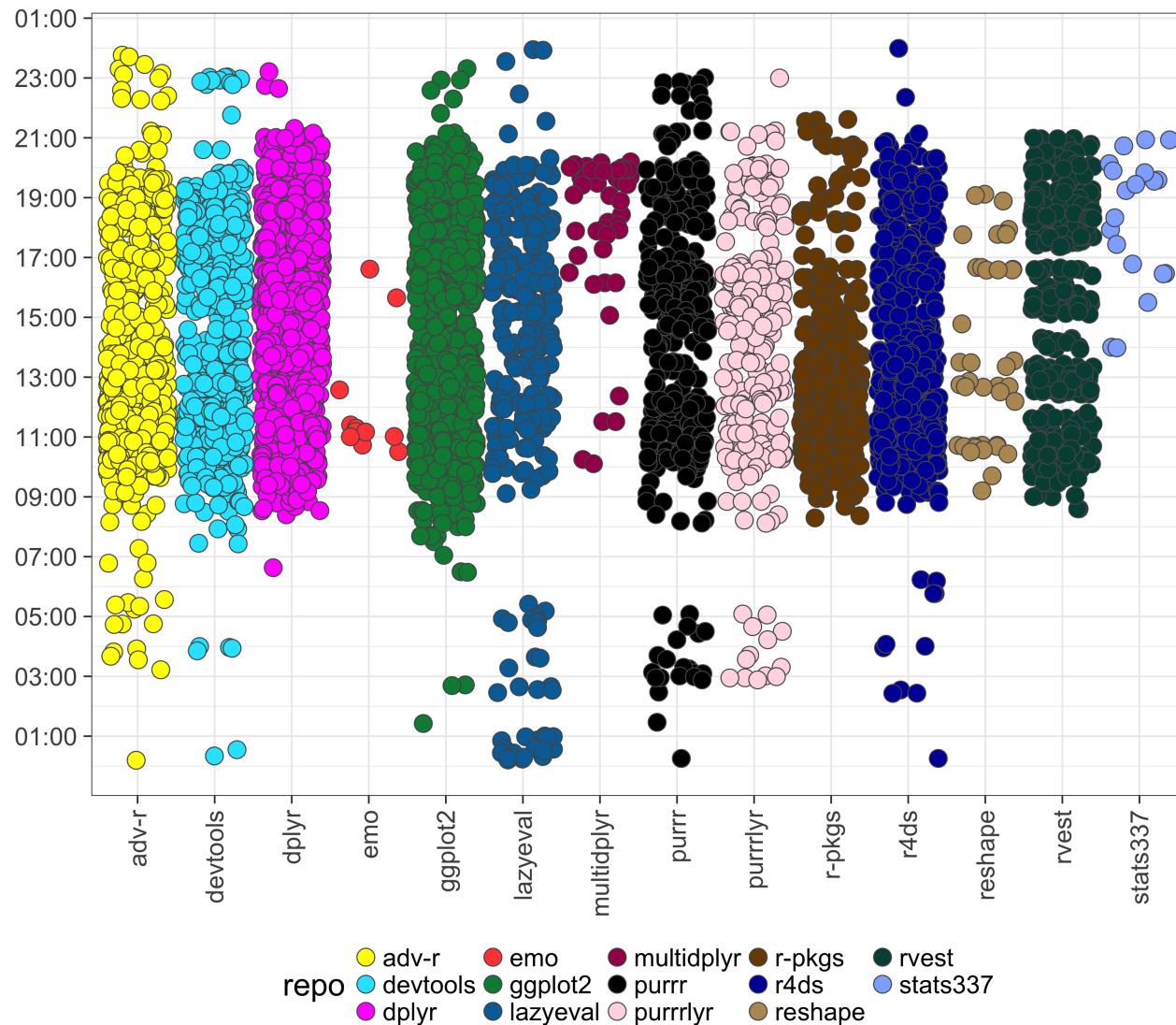
Si miras adv, purr y
purrly su semana
laboral tiene 7 días



@hadleywickham



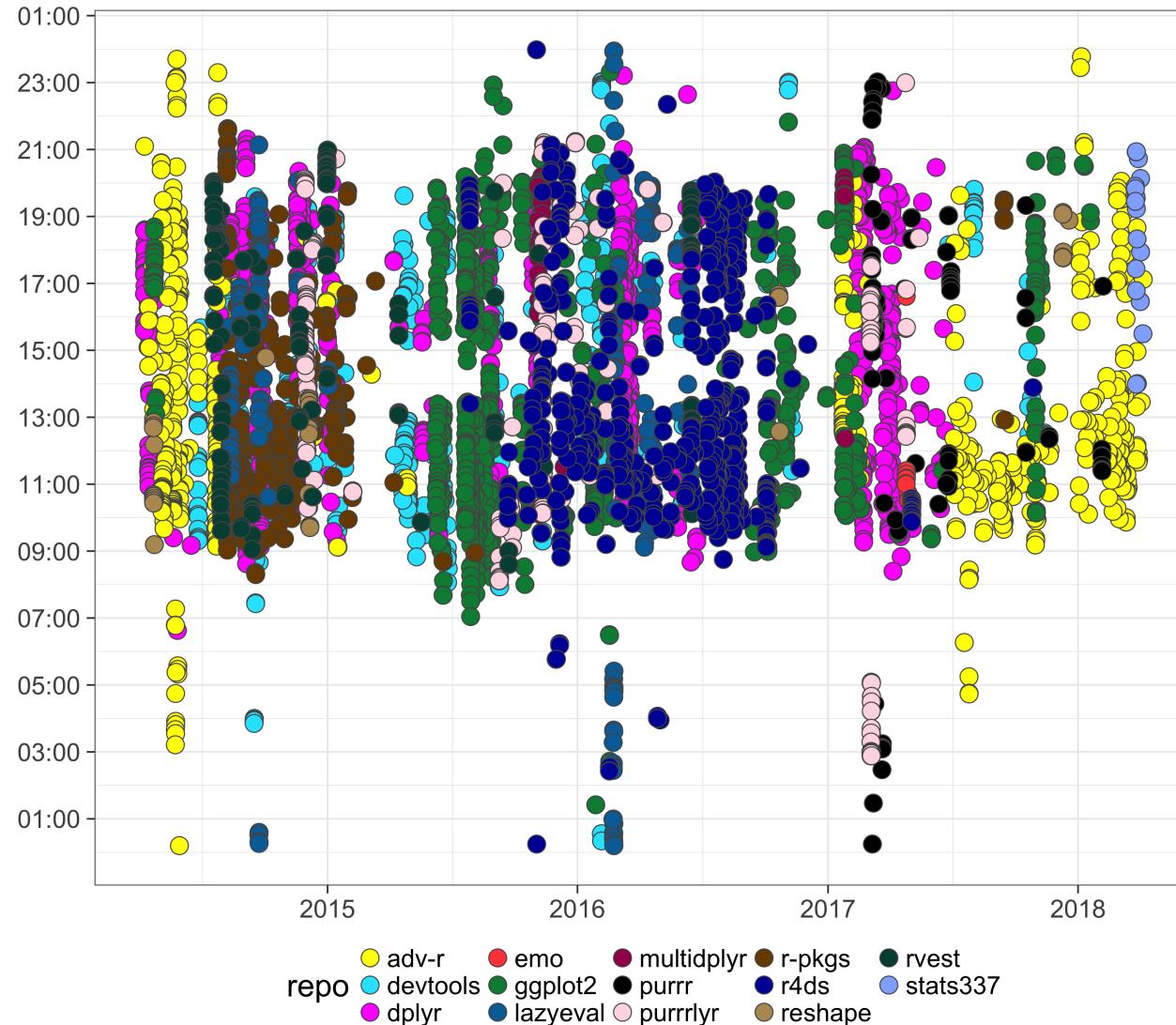
Que tan productivo es Hadley Wickham?



Trabaja de 8am a 21pm



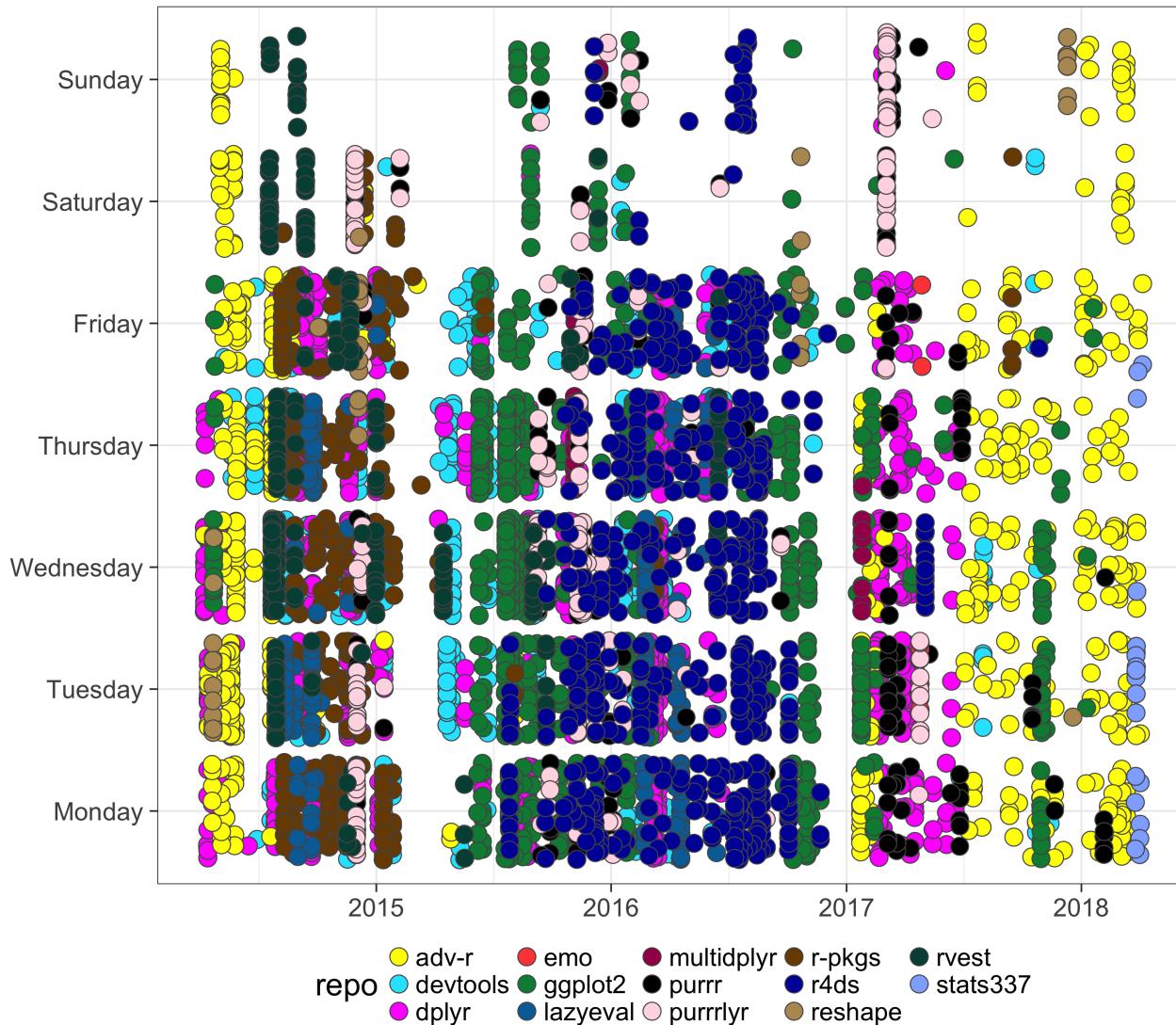
Que tan productivo es Hadley Wickham?



.... y hace varios años ...

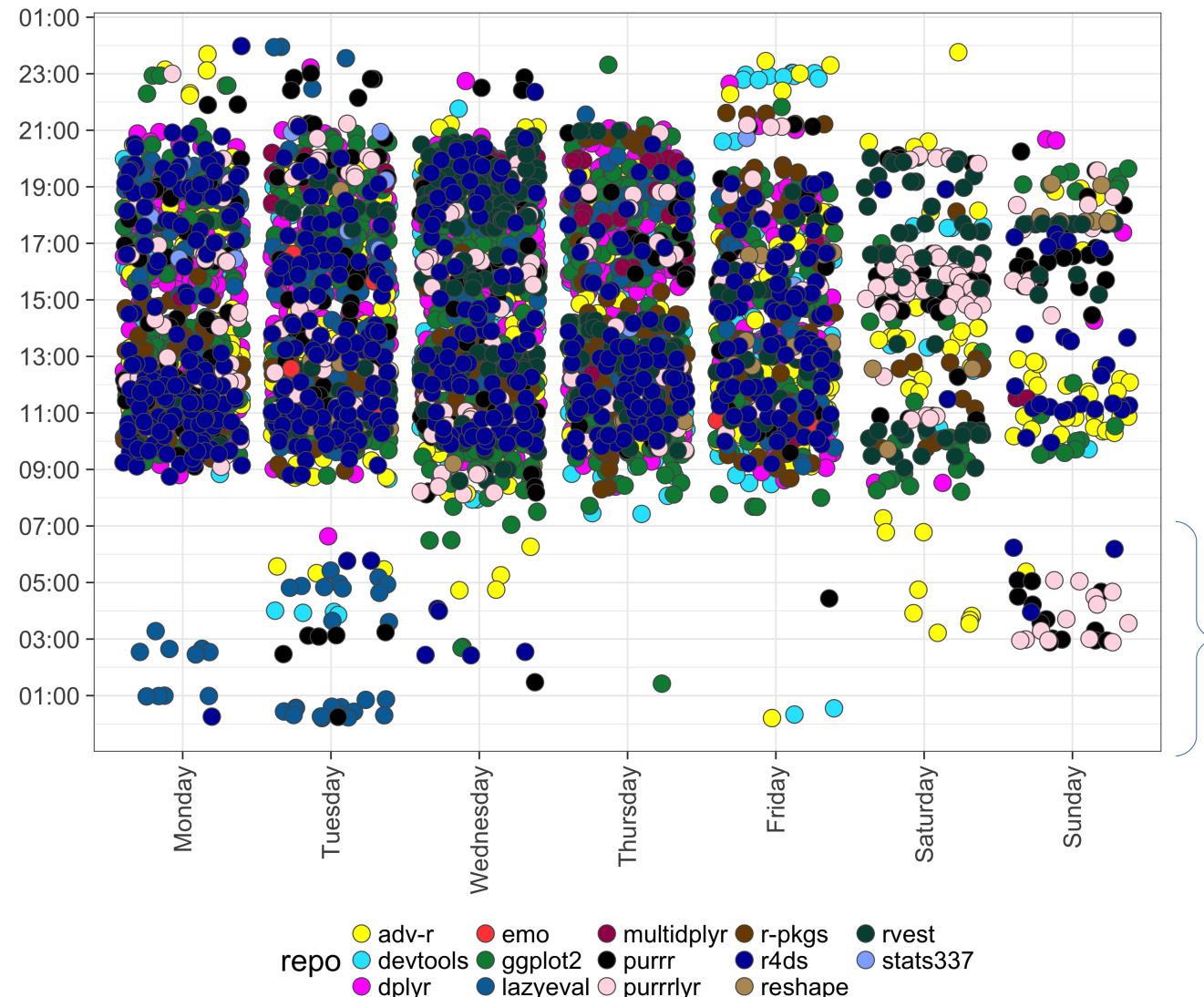


Que tan productivo es Hadley Wickham?





Que tan productivo es Hadley Wickham?



eFotosBebes.com



@hadleywickham



Es Hadley Wickham de otro planeta?



*“Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**”*



@hadleywickham



Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**"*

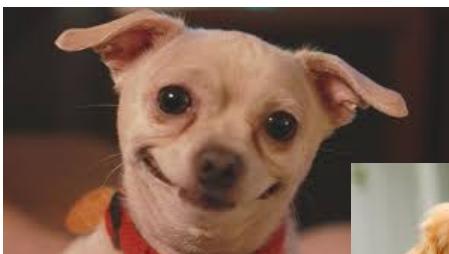




Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**"*





Es Hadley Wickham de otro planeta?



*“Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**”*



*“Outside of work, **I love to bake and make cocktails.** You might enjoy trying some of the recipes from my family recipes website.”*



Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**"*



*"Outside of work, **I love to bake and make cocktails**. You might enjoy trying some of the recipes from my family recipes website."*





Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**"*



*"Outside of work, **I love to bake and make cocktails**. You might enjoy trying some of the recipes from my family recipes website."*





Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand but I currently live in Houston, TX **with my partner and two dogs.**"*



*"Outside of work, **I love to bake and make cocktails**. You might enjoy trying some of the recipes from my family recipes website."*





Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand currently live in Houston with my partner and two dogs."*



*"Outside of work, **I love to bake and make cocktails**. You might enjoy trying some of the recipes from my family recipes website."*





Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley Wickham, **Chief Scientist at RStudio**, and an Adjunct Professor of Statistics at the University of Auckland, Stanford University, and Rice University. I build tools (computational and cognitive) that make data science easier, faster, and more fun. I'm from New Zealand currently live in Houston with my partner and two dogs."*



*"Outside of work, **I love to bake and make cocktails**. You might enjoy trying some of the recipes from my family recipes website."*





Es Hadley Wickham de otro planeta?



"Hi! I'm Hadley Wickham, Scientist at RStudio. Professor of Statistics at the University of Auckland, Stanford University, Rice University. I work on computational and statistical methods to make data science more fun. I'm from New Zealand, currently live in Edinburgh, Scotland with my partner and two dogs."



amazon.es Prueba Prime

Hola. Elige tu dirección Todos los departamentos Mi Amazon.es Ofertas Cheques regalo Vender Ayuda Cuenta y listas Pedidos Suscríbete a Prime Cesta

Libros Todos los géneros Búsqueda avanzada Novedades Los más vendidos Libros de texto Inglés Otros idiomas Infantiles y juveniles Libros universitarios y de estudios superiores

Hadley Wickham



+ Seguir



46,91 € Tapa blanda 173,90 € Tapa blanda 124,46 € Tapa blanda 173,90 € Tapa blanda 33,28 € Tapa blanda 56,23 € Tapa blanda

Libros de Hadley Wickham

Mostrando 4 resultados Libros : Búsqueda avanzada Ordenar por Popularidad

Libros en Español Libros en Inglés Tapa blanda



[(R Packages)] [By (author) Hadley Wickham] published on (April, 2015) 30 abril 2015
Tapa blanda EUR 46,91
Más opciones de compra EUR 46,90 usado y nuevo (3 ofertas)

★★★★★ ▾ 1

 @hadleywickham

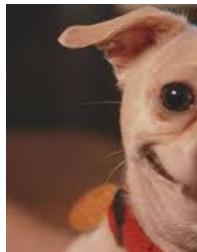


Es Hadley Wickham de otro planeta?



*"Hi! I'm Hadley W
Scientist at RStu*

*Profes
of Au
Rice l
(com
make
more
curre
partn*



amazon.es Prueba Prime

Hola. Identifícate Cuenta y listas Pedidos Suscríbete a Prime 0 Cesta

Libros Todos los resultados

Infantiles y juveniles Libros universitarios y de estudios superiores

Series

Hadley Wickham	ggplot2	O'REILLY'	Advanced R
Hadley Wickham	Hadley Wickham	R for Data Science	Hadley Wickham
ggplot2	Bright Graphics for Data Analysis	IMPORT, TRANSFORM, VISUALIZE, AND MODEL DATA	Advanced R
Springer	O'Reilly	Hadley Wickham & Garrett Grolemund	CRC Press
173,90 €	33,28 €	56,23 €	56,23 €
Tapa blanda	Tapa blanda	Tapa blanda	Tapa blanda

Ordenar por Popularidad

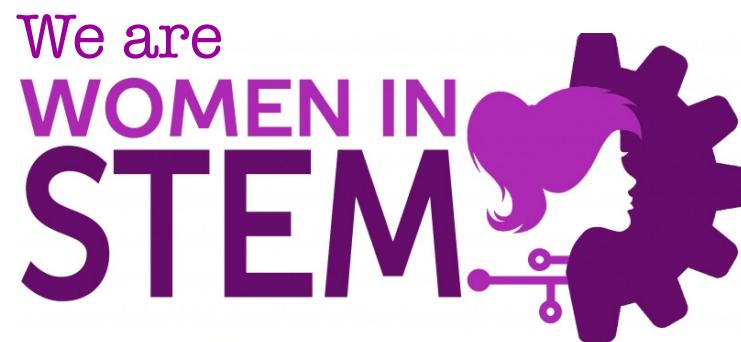
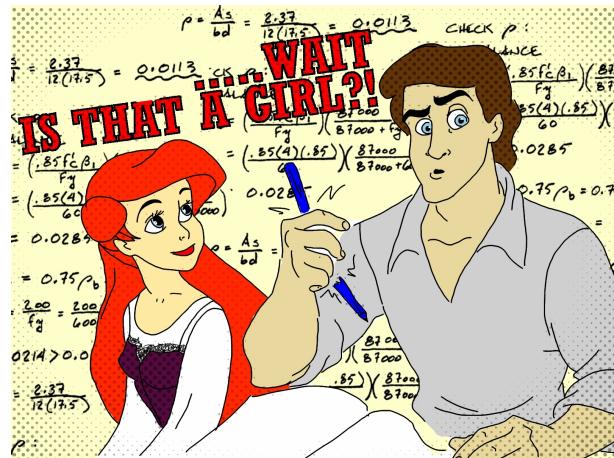
adley Wickham] published on (April, 2015) 30 abril 2015



@hadleywickham



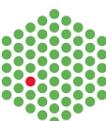
MUCHAS GRACIAS!!



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Institut Pasteur



EMBL-EBI