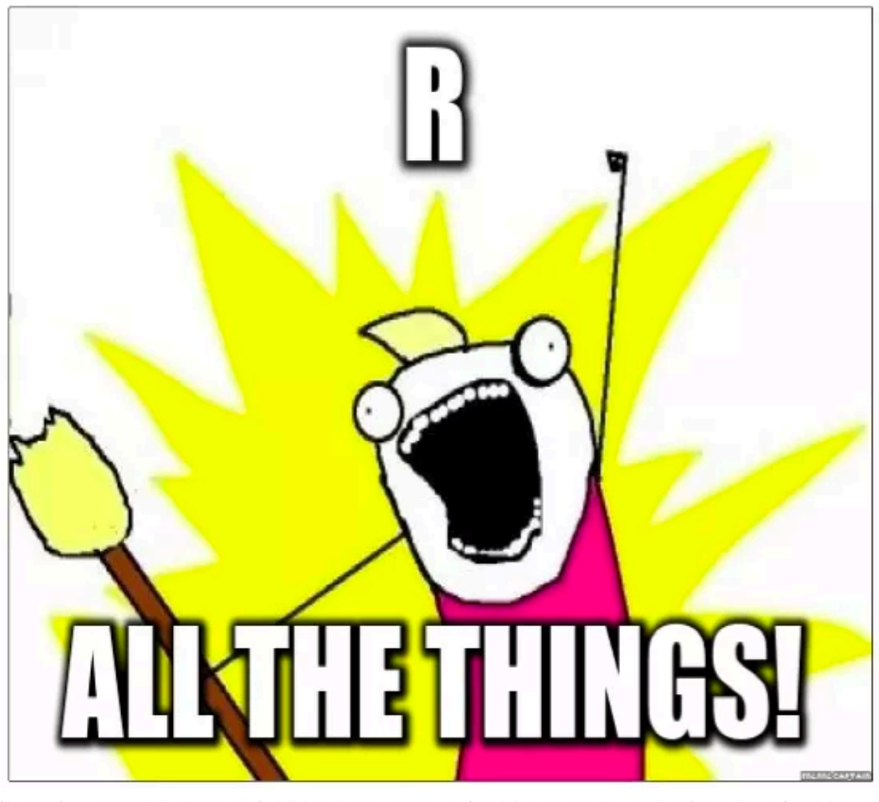
```
@Seminar{,
        title = {Levelling up R for Statistical Research and
Teaching},
        author = {Aleeza Gerstein},
        year = {2019},
        note = {UManitoba Statistics Department},
        url = {github.com/acgerstein/seminars/
RAllTheThings-AleezaGerstein.pdf},
}
```

@acgerstein

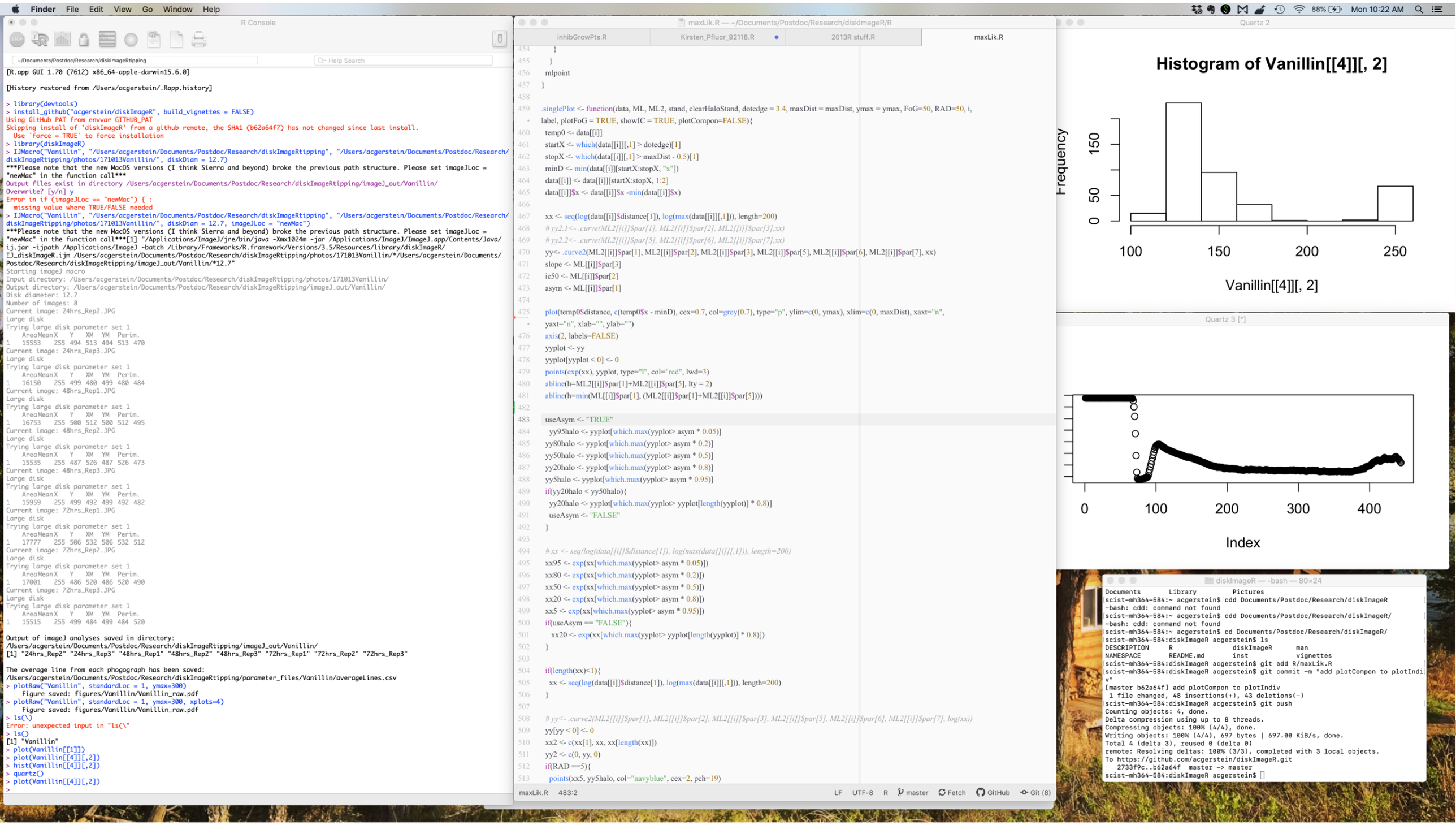


University of Manitoba

faculty of SCIENCE
discover the unknown + invent the future

```
install.packages("meme")
```

This is how I R (since 2004)



This is how I R (since 2004)

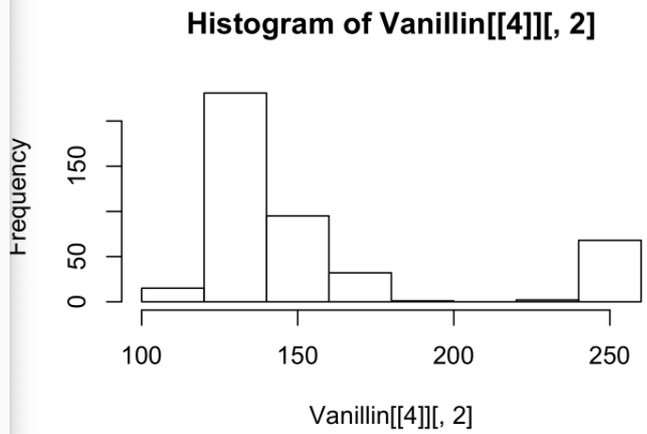
Base console

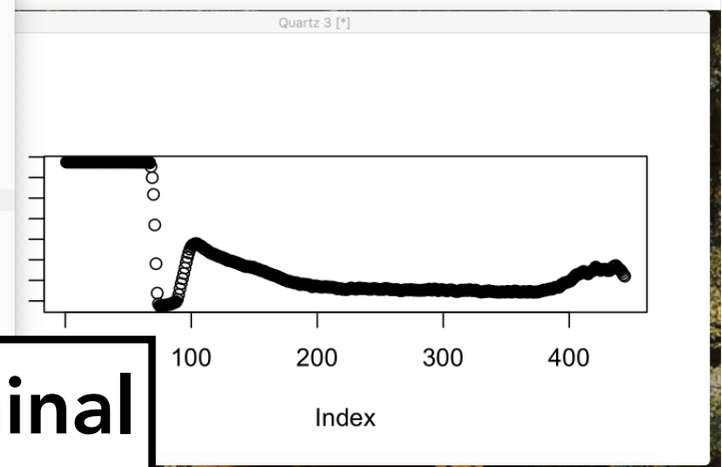
Base plot

Git terminal

```
R Console
~/Documents/Postdoc/Research/diskImageRtipping
[R.app GUI 1.70 (7612) x86_64-apple-darwin15.6.0]

[History restored from /Users/acgerstein/.Rapp.history]

> library(devtools)
> install_github("acgerstein/diskImageR", build_vignettes = FALSE)
Using GitHub PAT from envvar GITHUB_PAT
Skipping install of 'diskImageR' from a github remote, the SHA1 (b62a64f7) has not changed since last install.
  Use `force = TRUE` to force installation
> library(diskImageR)
> IJMacro("Vanillin", "/Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping", "/Users/acgerstein/Documents/Postdoc/Research/
diskImageRtipping/photos/171013Vanillin/", diskDiam = 12.7)
***Please note that the new MacOS versions (I think Sierra and beyond) broke the previous path structure. Please set imageJLoc =
"newMac" in the function call***
Output files exist in directory /Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/imageJ_out/Vanillin/
Overwrite? [y/n] y
Error in if (imageJLoc == "newMac") { :
  missing value where TRUE/FALSE needed
> IJMacro("Vanillin", "/Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping", "/Users/acgerstein/Documents/Postdoc/Research/
diskImageRtipping/photos/171013Vanillin/", diskDiam = 12.7, imageJLoc = "newMac")
***Please note that the new MacOS versions (I think Sierra and beyond) broke the previous path structure. Please set imageJLoc =
"newMac" in the function call***[1] "/Applications/ImageJ/jre/bin/java -Xmx1024m -jar /Applications/ImageJ/ImageJ.app/Contents/Java/
ij.jar -ijpath /Applications/ImageJ -batch /Library/Frameworks/R.framework/Versions/3.5/Resources/library/diskImageR/
IJ_diskImageR.ijm /Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/photos/171013Vanillin/*/Users/acgerstein/Documents/
Postdoc/Research/diskImageRtipping/imageJ_out/Vanillin/*12.7"
Starting imageJ macro
Input directory: /Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/photos/171013Vanillin/
Output directory: /Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/imageJ_out/Vanillin/
Disk diameter: 12.7
Number of images: 8
Current image: 24hrs_Rep2.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   15553   255 494 513 494 513 470
Current image: 24hrs_Rep3.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   16150   255 499 480 499 480 484
Current image: 48hrs_Rep1.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   16753   255 500 512 500 512 495
Current image: 48hrs_Rep2.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   15535   255 487 526 487 526 473
Current image: 48hrs_Rep3.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   15959   255 499 492 499 492 482
Current image: 72hrs_Rep1.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   17777   255 506 532 506 532 512
Current image: 72hrs_Rep2.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   17001   255 486 520 486 520 490
Current image: 72hrs_Rep3.JPG
Large disk
Trying large disk parameter set 1
    AreaMeanX   Y   XM  YM  Perim.
1   15515   255 499 484 499 484 520

Output of imageJ analyses saved in directory:
/Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/imageJ_out/Vanillin/
[1] "24hrs_Rep2" "24hrs_Rep3" "48hrs_Rep1" "48hrs_Rep2" "48hrs_Rep3" "72hrs_Rep1" "72hrs_Rep2" "72hrs_Rep3"

The average line from each phogograph has been saved:
/Users/acgerstein/Documents/Postdoc/Research/diskImageRtipping/parameter_files/Vanillin/averageLines.csv
> plotRaw("Vanillin", standardLoc = 1, ymax=300)
    Figure saved: figures/Vanillin/Vanillin_raw.pdf
> plotRaw("Vanillin", standardLoc = 1, ymax=300, xplots=4)
    Figure saved: figures/Vanillin/Vanillin_raw.pdf
> ls(\)
Error: unexpected input in "ls(\"
> ls()
[1] "Vanillin"
> plot(Vanillin[[1]])
> plot(Vanillin[[4]][,2])
> hist(Vanillin[[4]][,2])
> quartz()
> plot(Vanillin[[4]][,2])
>
```

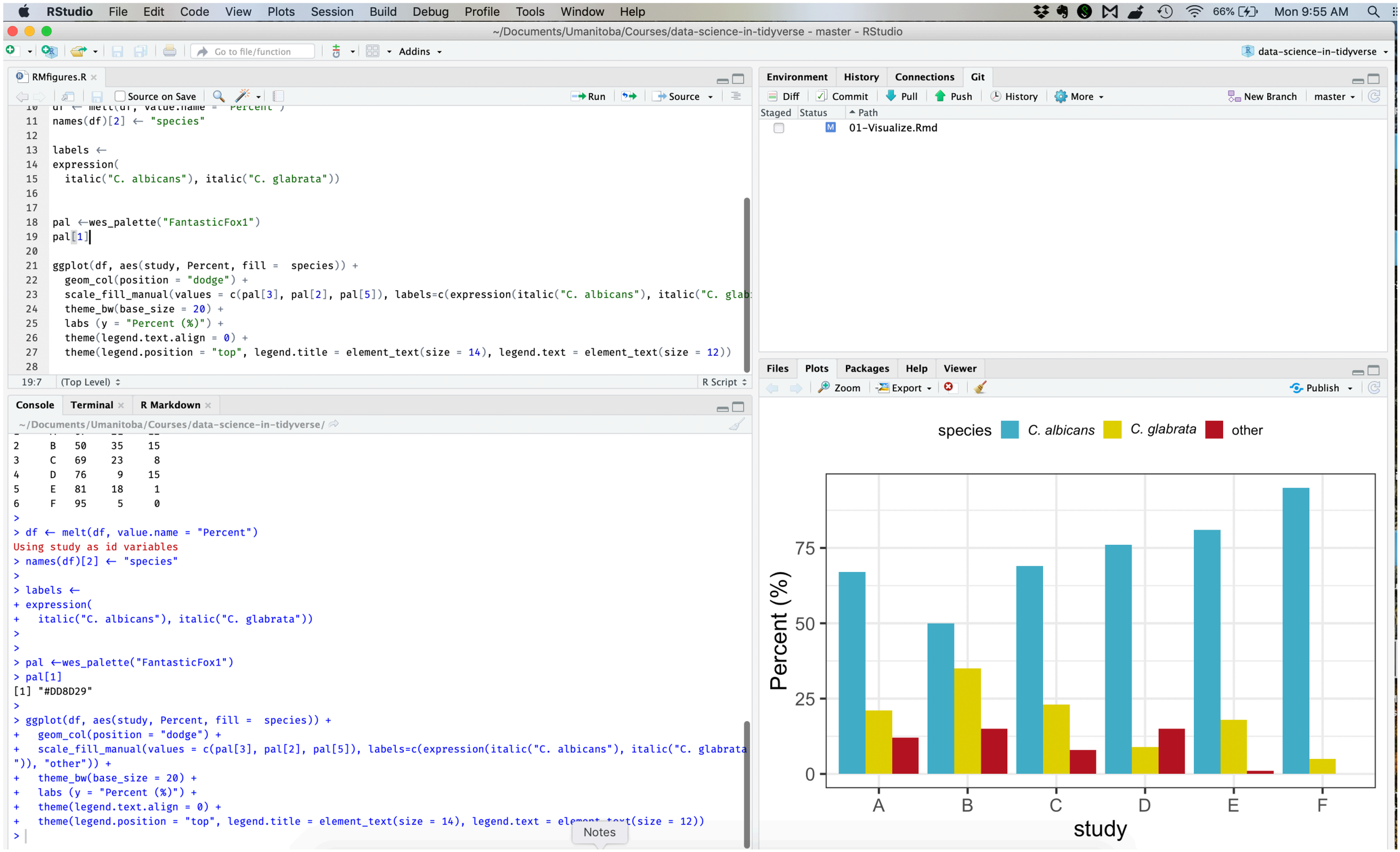
```
maxLik.R — ~/Documents/Postdoc/Research/diskImageR/R
inhibGrowPts.R | Kirsten_Pfluor_92118.R | 2013R stuff.R | maxLik.R

454     }
455     }
456   mlpoint
457   }
458
459   .singlePlot <- function(data, ML, ML2, stand, clearHaloStand, dotedge
      label, plotFoG = TRUE, showIC = TRUE, plotCompon=FALSE){
460     temp0 <- data[[i]]
461     startX <- which(data[[i]][,1] > dotedge)[1]
462     stopX <- which(data[[i]][,1] > maxDist - 0.5)[1]
463     minD <- min(data[[i]][startX:stopX, "x"])
464     data[[i]] <- data[[i]][startX:stopX, 1:2]
465     data[[i]]$x <- data[[i]]$x -min(data[[i]]$x)
466
467     xx <- seq(log(data[[i]]$distance[1]), log(max(data[[i]][,1])), length=2
468     # yy2.1<- .curve(ML2[[i]]$par[1], ML2[[i]]$par[2
469     # yy2.2<- .curve(ML2[[i]]$par[5], ML2[[i]]$par
470     yy<- .curve2(ML2[[i]]$par[1], ML2[[i]]$par[2], M                    2[[i]]$par[5], ML2[[i]]$par[6], ML2[[i]]$par[7], xx)
471     slope <- ML[[i]]$par[3]
472     ic50 <- ML[[i]]$par[2]
473     asym <- ML[[i]]$par[1]
474
475     plot(temp0$distance, c(temp0$x - minD), cex=0.7, col=grey(0.7), type="p", ylim=c(0, ymax), xlim=c(0, maxDist), xaxt="n",
          yaxt="n", xlab="", ylab="")
476     axis(2, labels=FALSE)
477     yyplot <- yy
478     yyplot[yyplot < 0] <- 0
479     points(exp(xx), yyplot, type="l", col="red", lwd=3)
480     abline(h=ML2[[i]]$par[1]+ML2[[i]]$par[5], lty = 2)
481     abline(h=min(ML[[i]]$par[1], (ML2[[i]]$par[1]+ML2[[i]]$par[5])))
482
483     useAsym <- "TRUE"
484       yy95halo <- yyplot[which.max(yyplot> asym * 0.05)]
485     yy80halo <- yyplot[which.max(yyplot> asym * 0.2)]
486     yy50halo <- yyplot[which.max(yyplot> asym * 0.5)]
487     yy20halo <- yyplot[which.max(yyplot> asym * 0.8)]
488     yy5halo <- yyplot[which.max(yyplot> asym * 0.95)]
489     if(yy20halo < yy50halo){
490       yy20halo <- yyplot[which.max(yyplot> yyplot[length(yyplot)] * 0.8)]
491       useAsym <- "FALSE"
492     }
493
494     # xx <- seq(log(data[[i]]$distance[1]), log(max(data[[i]][,1])), length=200)
495     xx95 <- exp(xx[which.max(yyplot> asym * 0.05)])
496     xx80 <- exp(xx[which.max(yyplot> asym * 0.2)])
497     xx50 <- exp(xx[which.max(yyplot> asym * 0.5)])
498     xx20 <- exp(xx[which.max(yyplot> asym * 0.8)])
499     xx5 <- exp(xx[which.max(yyplot> asym * 0.95)])
500     if(useAsym == "FALSE"){
501       xx20 <- exp(xx[which.max(yyplot> yyplot[length(yyplot)] * 0.8)])
502     }
503
504     if(length(xx)<1){
505       xx <- seq(log(data[[i]]$distance[1]), log(max(data[[i]][,1])), length=200)
506     }
507
508     # yy<- .curve2(ML2[[i]]$par[1], ML2[[i]]$par[2], ML2[[i]]$par[3], ML2[[i]]$par[5], ML2[[i]]$par[6], ML2[[i]]$par[7], log(xx))
509     yy[yy < 0] <- 0
510     xx2 <- c(xx[1], xx, xx[length(xx)])
511     yy2 <- c(0, yy, 0)
512     if(RAD ==5){
513       points(xx5, yy5halo, col="navyblue", cex=2, pch=19)

maxLik.R  483:2   LF  UTF-8  R  master  Fetch  GitHub  Git (8)
```



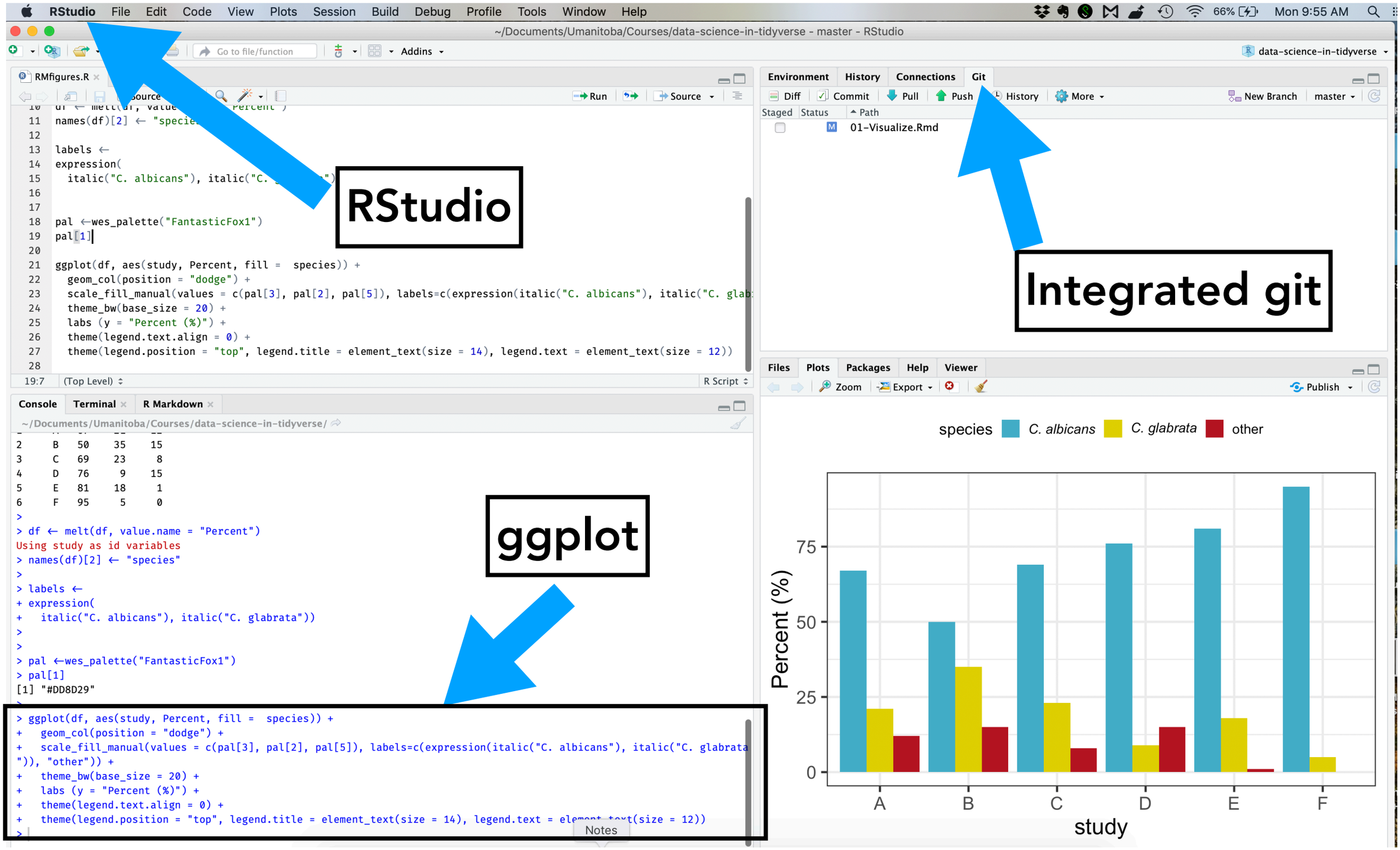



```
diskImageR — -bash — 80×24
Documents       Library         Pictures
(scist-mh364-584:~ acgerstein$ cdd Documents/Postdoc/Research/diskImageR
-bash: cdd: command not found
(scist-mh364-584:~ acgerstein$ cdd Documents/Postdoc/Research/diskImageR/
-bash: cdd: command not found
(scist-mh364-584:~ acgerstein$ cd Documents/Postdoc/Research/diskImageR/
(scist-mh364-584:diskImageR acgerstein$ ls
DESCRIPTION     R               diskImageR      man
NAMESPACE       README.md       inst            vignettes
(scist-mh364-584:diskImageR acgerstein$ git add R/maxLik.R
(scist-mh364-584:diskImageR acgerstein$ git commit -m "add plotCompon to plotIndi
v"
[master b62a64f] add plotCompon to plotIndiv
 1 file changed, 48 insertions(+), 43 deletions(-)
(scist-mh364-584:diskImageR acgerstein$ git push
Counting objects: 4, done.
Delta compression using up to 8 threads.
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 697 bytes | 697.00 KiB/s, done.
Total 4 (delta 3), reused 0 (delta 0)
remote: Resolving deltas: 100% (3/3), completed with 3 local objects.
To https://github.com/acgerstein/diskImageR.git
   2733f9c..b62a64f  master -> master
scist-mh364-584:diskImageR acgerstein$
```

This is how my students R



This is how my students R



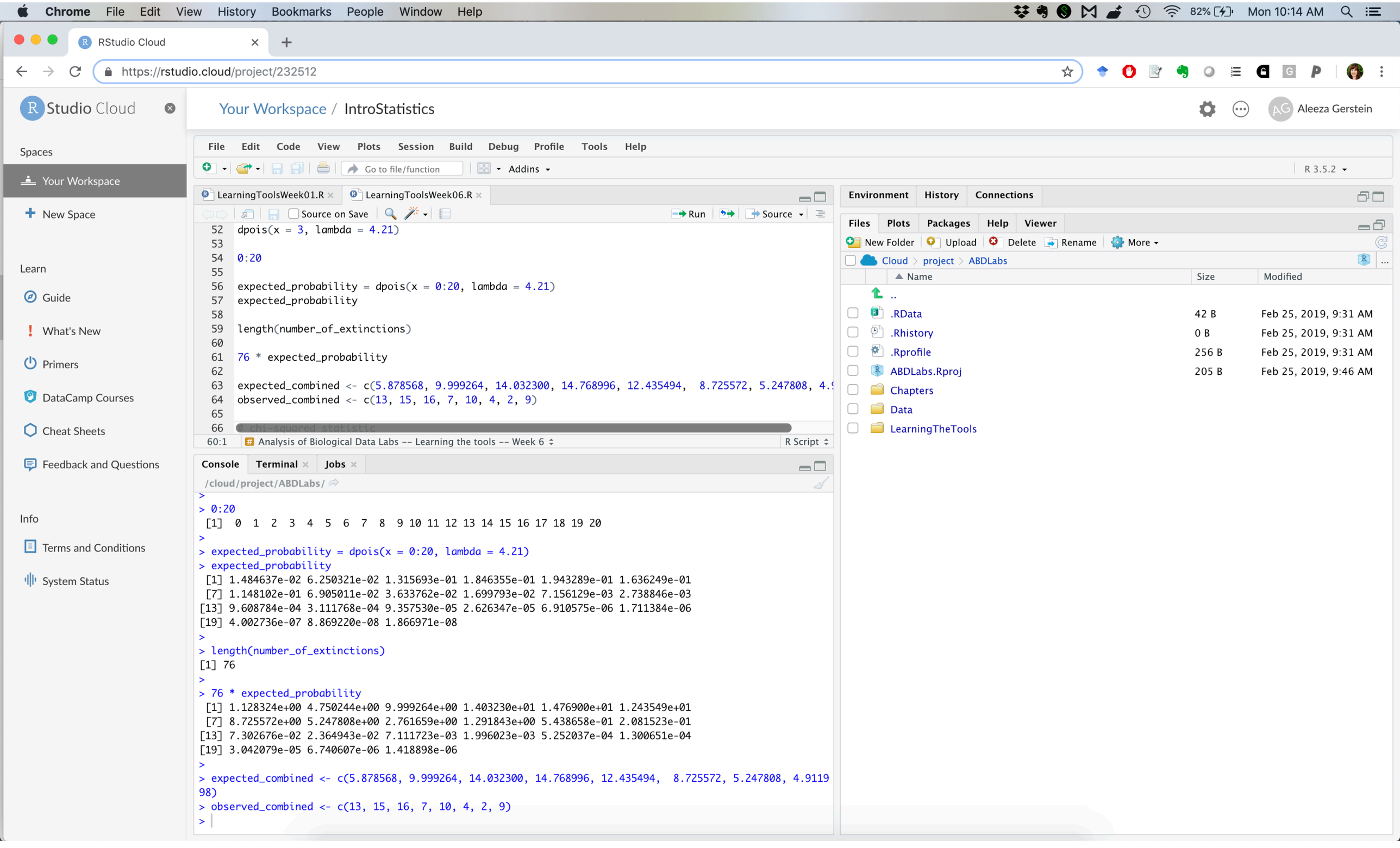
RStudio (the IDE & company) is the future

R: Engine	RStudio: Dashboard
	

http://moderndive.com/

Chester Ismay & Albert Y. Kim: Modern Dive

This is how we could teach R



This is how we could teach R

RStudio Cloud through the web!

Chrome File Edit View History Bookmarks People Window Help — 82% — Mon 10:14 AM

RStudio Cloud

https://rstudio.cloud/project/232512

Studio Cloud

Your Workspace / IntroSta cs

Aleeza Gerstein

Spaces
- Your Workspace
- New Space

Learn
- Guide
- What's New
- Primers
- DataCamp Courses
- Cheat Sheets
- Feedback and Questions

Info
- Terms and Conditions
- System Status

File Edit Code View Plots Session Build Debug Profile Help

Go to file/function — Addins — R 3.5.2

LearningToolsWeek01.R | LearningToolsWeek06.R

Source on Save

```
52 dpois(x = 3, lambda = 4.21)
53
54 0:20
55
56 expected_probability = dpois(x = 0:20, lambda = 4.21)
57 expected_probability
58
59 length(number_of_extinctions)
60
61 76 * expected_probability
62
63 expected_combined <- c(5.878568, 9.999264, 14.032300, 14.768996, 12.435494,  8.725572, 5.247808, 4.9
64 observed_combined <- c(13, 15, 16, 7, 10, 4, 2, 9)
65
66 # chi-squared statistic
```

60:1 — Analysis of Biological Data Labs -- Learning the tools -- Week 6 — R Script

Console | Terminal | Jobs

/cloud/project/ABDLabs/

```
> 0:20
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
>
> expected_probability = dpois(x = 0:20, lambda = 4.21)
> expected_probability
 [1] 1.484637e-02 6.250321e-02 1.315693e-01 1.846355e-01 1.943289e-01 1.636249e-01
 [7] 1.148102e-01 6.905011e-02 3.633762e-02 1.699793e-02 7.156129e-03 2.738846e-03
[13] 9.608784e-04 3.111768e-04 9.357530e-05 2.626347e-05 6.910575e-06 1.711384e-06
[19] 4.002736e-07 8.869220e-08 1.866971e-08
>
> length(number_of_extinctions)
[1] 76
>
> 76 * expected_probability
 [1] 1.128324e+00 4.750244e+00 9.999264e+00 1.403230e+01 1.476900e+01 1.243549e+01
 [7] 8.725572e+00 5.247808e+00 2.761659e+00 1.291843e+00 5.438658e-01 2.081523e-01
[13] 7.302676e-02 2.364943e-02 7.111723e-03 1.996023e-03 5.252037e-04 1.300651e-04
[19] 3.042079e-05 6.740607e-06 1.418898e-06
>
> expected_combined <- c(5.878568, 9.999264, 14.032300, 14.768996, 12.435494,  8.725572, 5.247808, 4.9119
98)
> observed_combined <- c(13, 15, 16, 7, 10, 4, 2, 9)
>
```

Cloud > project > ABDLabs

Name	Size	Modified
..		
.RData	42 B	Feb 25, 2019, 9:31 AM
.Rhistory	0 B	Feb 25, 2019, 9:31 AM
.Rprofile	256 B	Feb 25, 2019, 9:31 AM
ABDLabs.Rproj	205 B	Feb 25, 2019, 9:46 AM
Chapters		
Data		
LearningTheTools		

A very brief history

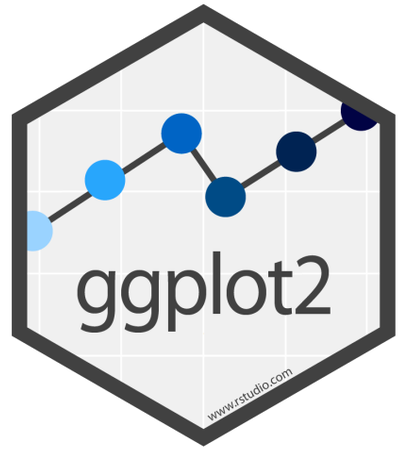
S → 

1976
Rick Becker and John Chambers

1996
Ross Ihaka and Robert Gentleman
[R Foundation in 2003]

Carlos Grajales: How R took the world of statistics by storm

A very brief history

S →  → 

1976
Rick Becker and John Chambers

1996
Ross Ihaka and Robert Gentleman
[R Foundation in 2003]

2005
Hadley Wickham

A very brief history

S → R → dplyr

1976
Rick Becker and John Chambers

Bell Laboratories

1996
Ross Ihaka and Robert Gentleman
[R Foundation in 2003]

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

2013
Hadley Wickham

A very brief history

S → R → tidyverse

1976
Rick Becker and John Chambers

Bell Laboratories

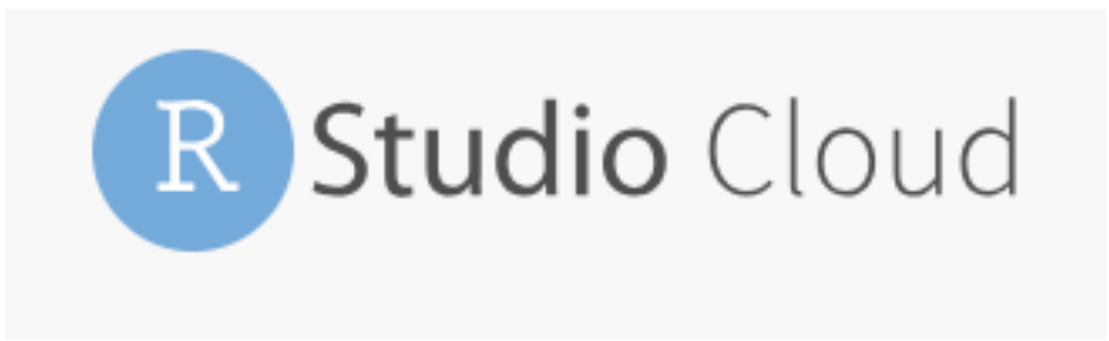
1996
Ross Ihaka and Robert Gentleman
[R Foundation in 2003]

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

2015
RStudio

Talk outline

vs.



- software environment for statistical computing and graphics.

- flexible and programmable

- total control to user

Roger Peng: https://simplystatistics.org/2018/07/12/use-r-keynote-2018/

packages for everything

car – car's Anova function is popular for making type II and type III Anova tables.

mgcv – Generalized Additive Models

lme4/nlme – Linear and Non-linear mixed effects models

randomForest – Random forest methods from machine learning

multcomp – Tools for multiple comparison testing

vcd – Visualization tools and tests for categorical data

glmnet – Lasso and elastic-net regression methods with cross validation

survival – Tools for survival analysis

caret – Tools for training regression and classification models

Hosted on CRAN, bioconductor, GitHub

https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages

vs.

- software environment for statistical computing and graphics.
- flexible and programmable
- total control to user

- opinionated collection of R packages designed for data science ("metapackage")
- designed for users with no previous programming experience, encountering data science and statistics for the first time
- guides users through workflows that improve reproducibility and communication

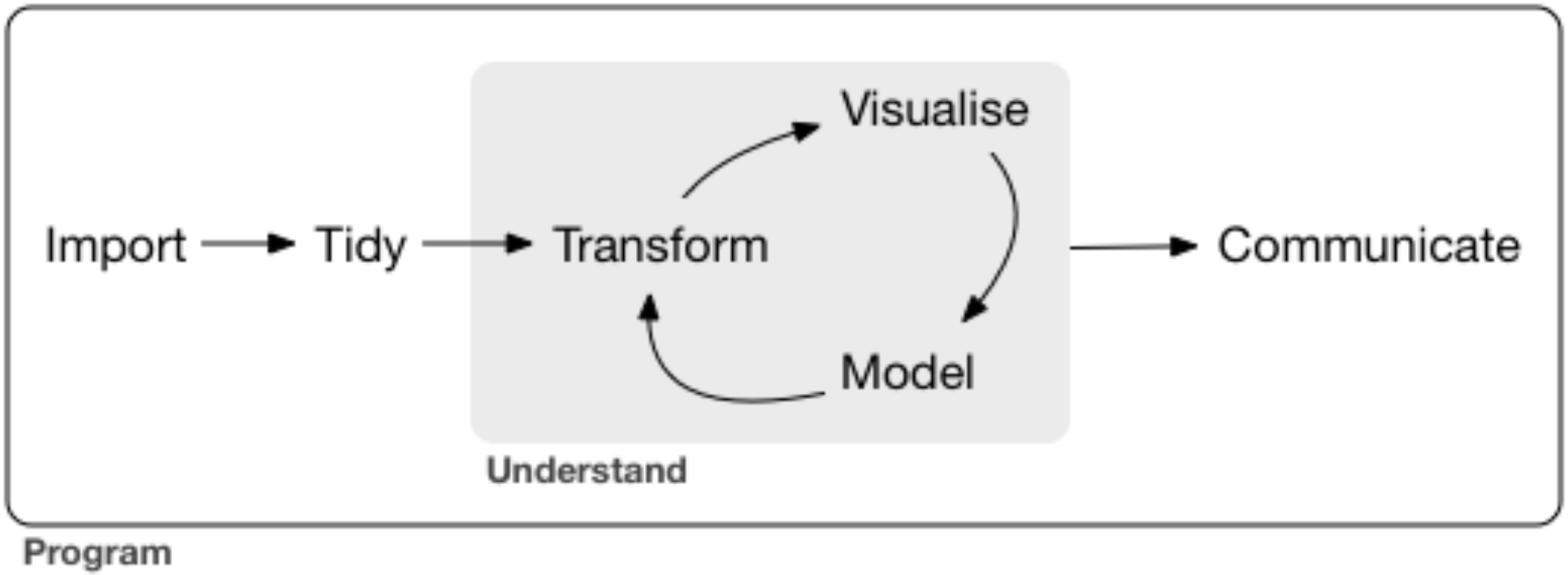
Roger Peng: https://simplystatistics.org/2018/07/12/use-r-keynote-2018/

```
aggregate(airquality[, “Ozone”],
          list(Month = airquality[, “Month”]),
          mean, na.rm = TRUE)
```


```
group_by(airquality, Month) %>%
   summarize(o3 = mean(Ozone, na.rm = TRUE))
```

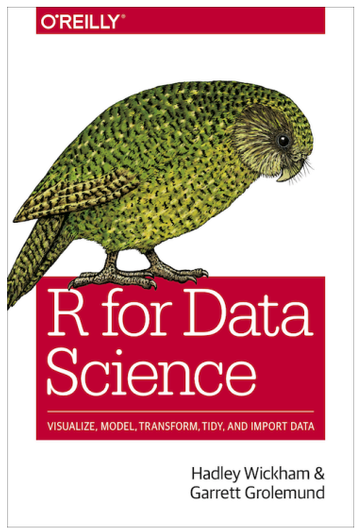
Roger Peng: https://simplystatistics.org/2018/07/12/use-r-keynote-2018/

Tidyverse workflow



Hadley Whickham: R for Data Science

Tidy Data. Journal of Statistical Software, Vol. 59, No. 10.



Consistency

- **Variables, functions, operators follow regular patterns and syntax**
  - First argument of a function is always a tidy data frame

baker	cinnamon_1	cardamom_2	nutmeg_3
Emma	1	0	1
Harry	1	1	1
Ruby	1	0	1
Zainab	0	NA	0

baker	spice	correct
Emma	cinnamon_1	1
Harry	cinnamon_1	1
Ruby	cinnamon_1	1
Zainab	cinnamon_1	0
Emma	cardamom_2	0
Harry	cardamom_2	1
Ruby	cardamom_2	0
Zainab	cardamom_2	NA
Emma	nutmeg_3	1
Harry	nutmeg_3	1
Ruby	nutmeg_3	1
Zainab	nutmeg_3	0

Alison Hill
@rstudio

H. Wickham. Reshaping data with the reshape package. Journal of Statistical Software, 21 (12), 2007c. URL http://www.jstatsoft.org/v21/i12/paper.

Readability

- **Pipe allows operations coded the way you read**
  - no nested function calls
  - no intermediate steps
  - minimize creation of local variables and overwriting original data

https://www.r-bloggers.com/readable-code-with-pipes/

Readability

- Pipe allows operations coded the way you read

```
> head(ecom2)
  referrer n_pages duration purchase
1   google       1      693    false
2    yahoo       1      459    false
3   direct       1      996    false
4     bing      18      468     true
5    yahoo       1      955    false
6    yahoo       5      135    false
```

```
> y <- sqrt(ecom2$n_pages)
```

```
> y <-
+   ecom2$n_pages %>%
+   sqrt()
```

https://www.r-bloggers.com/readable-code-with-pipes/

Readability

- **Pipe allows operations coded the way you read**

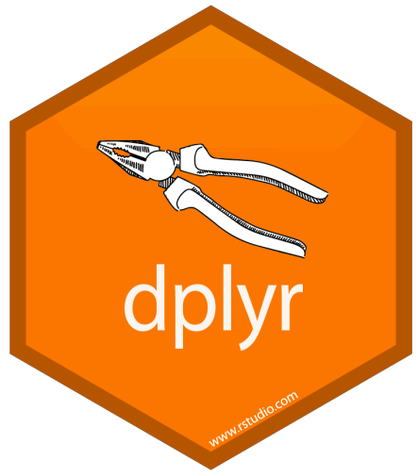
```
> head(ecom)
  referrer n_pages duration purchase
1   google       1      693    false
2    yahoo       1      459    false
3   direct       1      996    false
4     bing      18      468     true
5    yahoo       1      955    false
6    yahoo       5      135    false
```

```
> ecom1 <- subset(ecom, purchase)
> cor(ecom1$n_pages, ecom1$duration)
[1] 0.4290905
```

```
> ecom %>%
+   subset(purchase) %$%
+   cor(n_pages, duration)
[1] 0.4290905
```

https://www.r-bloggers.com/readable-code-with-pipes/

Readability



- **Function names are verbs**

```
# dplyr
crime.by.state <- read.csv("CrimeStatebyState.csv")
final <- crime.by.state %>%
            filter(State=="New York", Year==2005) %>%
            arrange(desc(Count)) %>%
            select(Type.of.Crime, Count) %>%
            mutate(Proportion=Count/sum(Count)) %>%
            group_by(Type.of.Crime) %>%
            summarise(num.types = n(), counts = sum(Count))
```

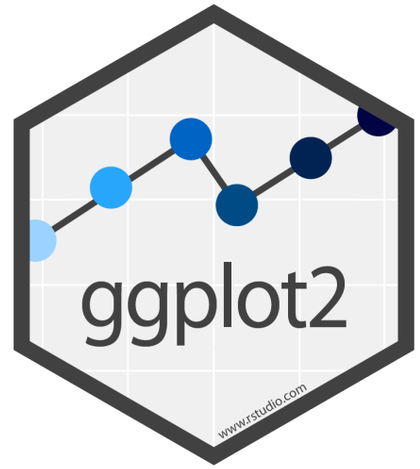
Tony Fischetti: How dplyr replaced my most common r idioms

Readability

```
# base R
crime.by.state <- read.csv("CrimeStatebyState.csv")
crime.ny.2005 <- crime.by.state[crime.by.state$Year==2005 &
                                crime.by.state$State=="New York",
                                c("Type.of.Crime", "Count")]
crime.ny.2005 <- crime.ny.2005[order(crime.ny.2005$Count,
                                     decreasing=TRUE), ]
crime.ny.2005$Proportion <- crime.ny.2005$Count /
                            sum(crime.ny.2005$Count)
summary1 <- aggregate(Count ~ Type.of.Crime,
                      data=crime.ny.2005,
                      FUN=sum)
summary2 <- aggregate(Count ~ Type.of.Crime,
                      data=crime.ny.2005,
                      FUN=length)
final <- merge(summary1, summary2,
               by="Type.of.Crime")
```

Tony Fischetti: How dplyr replaced my most common r idioms

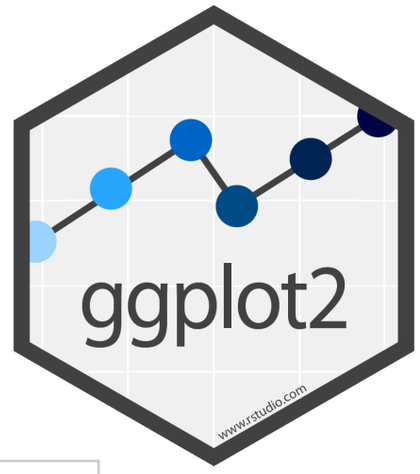
Philosophy of visualisation



"In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinates system."

Hadley Wickham (2010) A Layered Grammar of Graphics, Journal of Computational and Graphical Statistics, 19:1, 3-28, DOI: 10.1198/jcgs.2009.07098

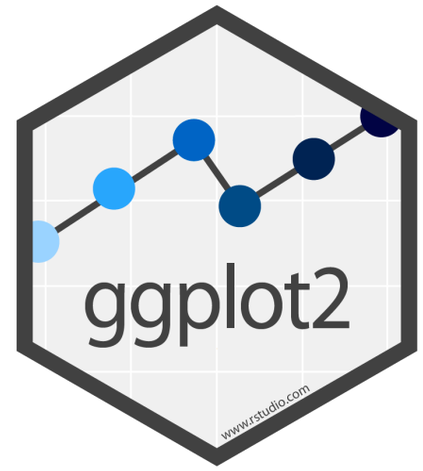
Philosophy of visualisation



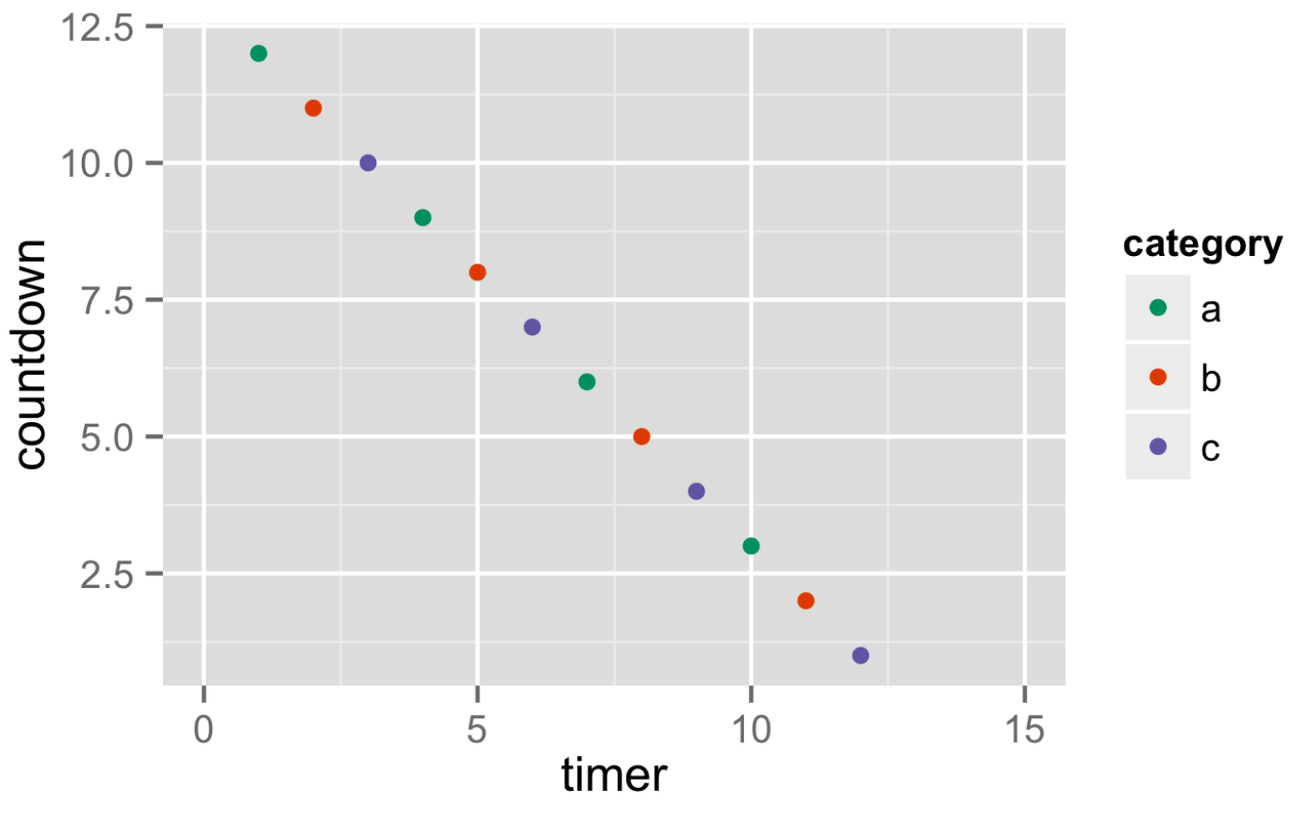
1	Data		The raw data that you want to plot
2	Geometries	`geom_`	The geometric shapes that will represent the data.
3	Aethetics	`aes()`	Aesthetics of the geometric and statistical objects, such as color, size, shape and position.
4	Scales	`scale_`	Maps between the data and the aesthetic dimensions, such as data range to plot width or factor values to colors
5	Statistical transformations	`stat_`	Statistical summaries of the data that can be plotted, such as quantiles, fitted curves (loess, linear models, etc.), sums and so on.
6	Coordinate systems	`coord_`	The transformation used for mapping data coordinates into the plane of the data rectangle.
7	Facets	`facet_`	The arrangement of the data into a grid of plots (also known as latticing, trellising or creating small multiples).
8	Visual Themes	`theme`	The overall visual defaults of a plot: background, grids, axe, default typeface, sizes, colors, etc.

https://www.r-bloggers.com/a-simple-introduction-to-the-graphing-philosophy-of-ggplot2/

Philosophy of visualisation

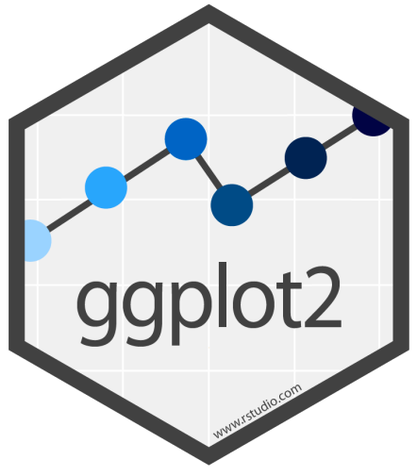



```
# Create some data for our example
some.data <- data.frame(timer = 1:12,
                        countdown = 12:1,
                        category = factor(letters[1:3]))

# Generate the plot
some.plot <- ggplot(data = some.data,
  aes(x = timer, y = countdown)) +
  geom_point(aes(colour = category)) +
  scale_x_continuous(limits = c(0, 15)) +
  scale_colour_brewer(palette = "Dark2") +
  coord_fixed(ratio=1)

# Display the plot
some.plot
```

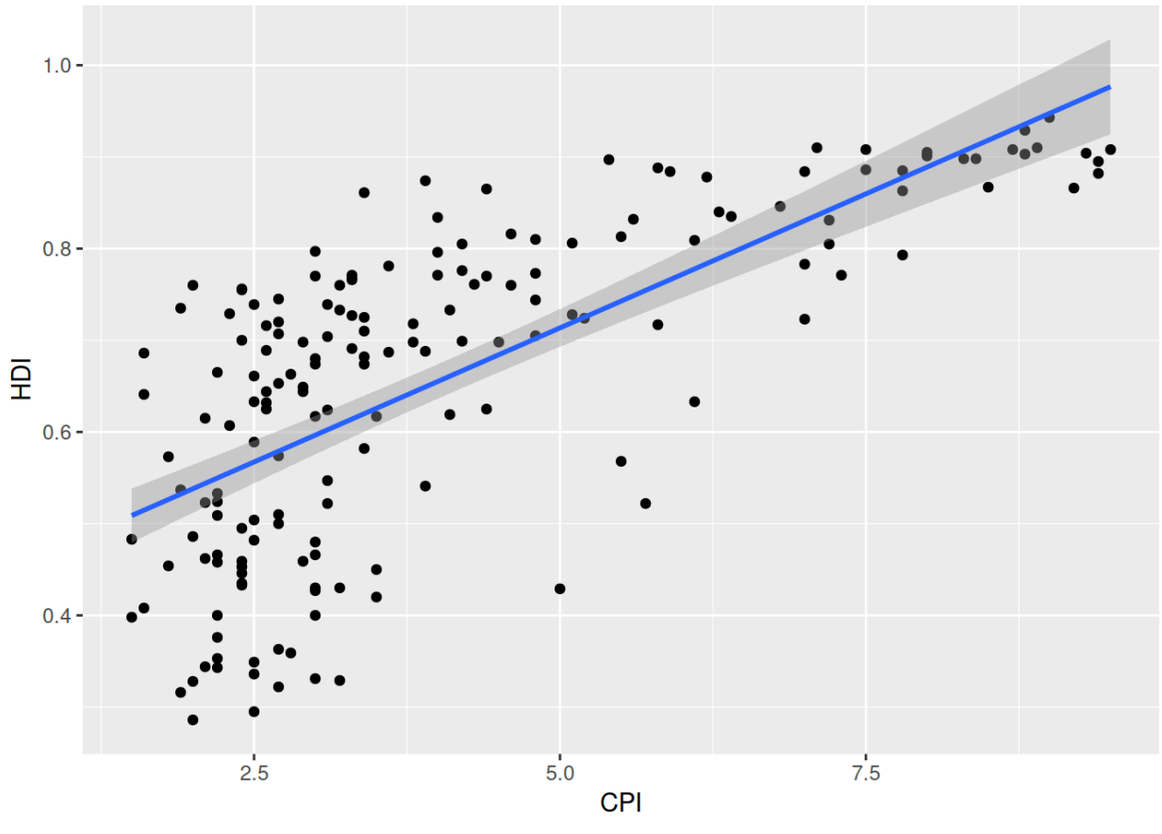


https://www.r-bloggers.com/a-simple-introduction-to-the-graphing-philosophy-of-ggplot2/

Philosophy of visualisation

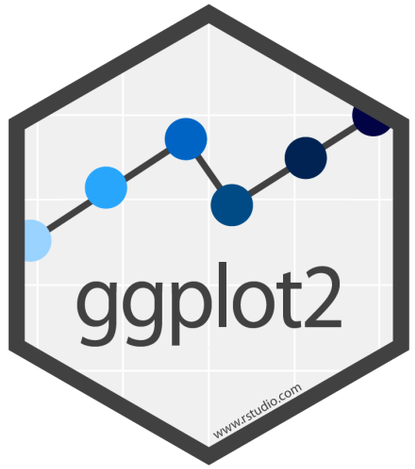



```
ggplot(dat, aes(x = CPI, y = HDI)) +
  geom_point() +
  geom_smooth(method = "lm")
```

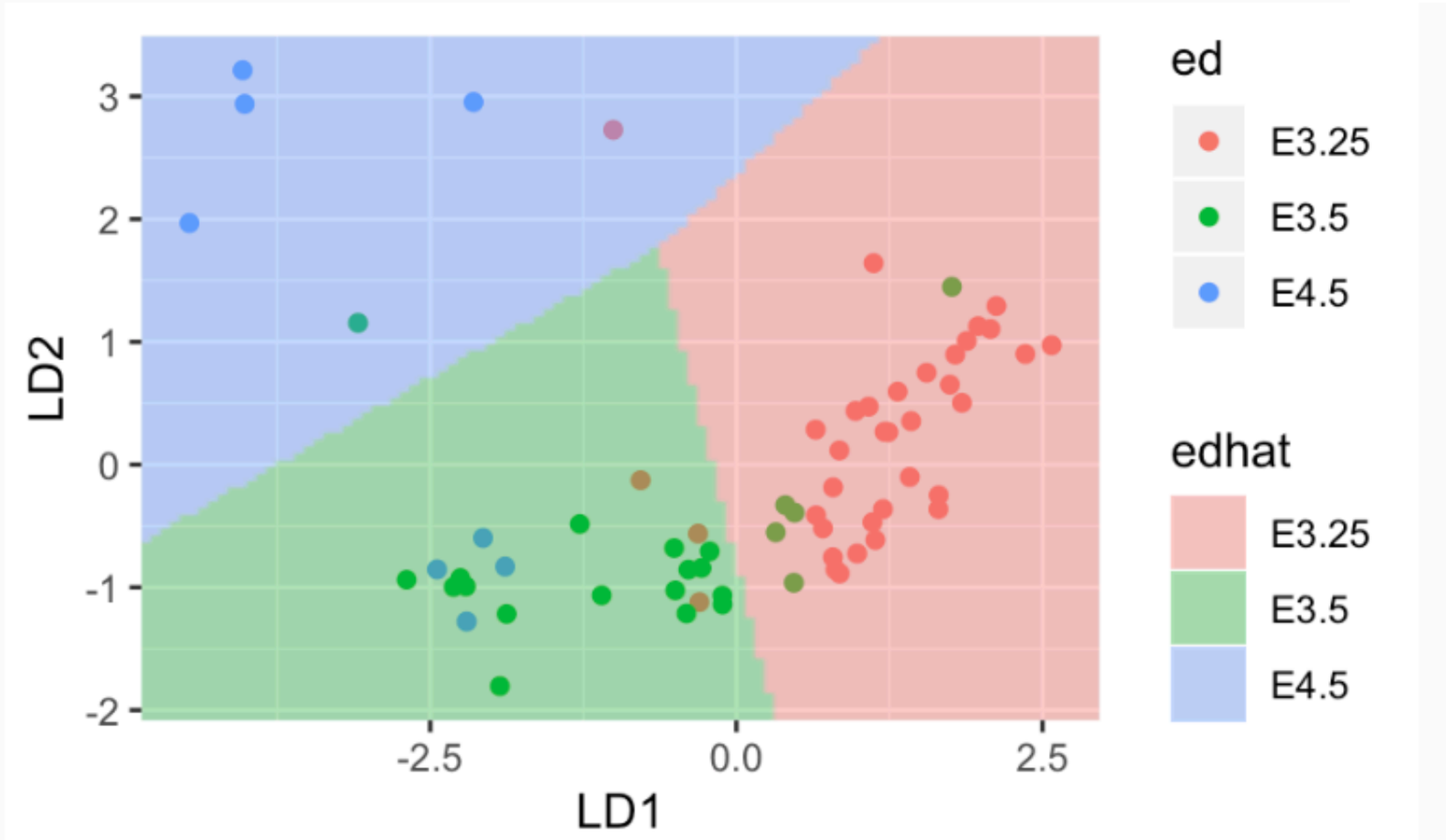


https://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html#geometric_objects_and_aesthetics

Philosophy of visualisation




```
ggplot() +
  geom_point(aes(x = LD1, y = LD2, colour = ed), data = ec_rot) +
  geom_raster(aes(x = LD1, y = LD2, fill = edhat),
              data = ec_grid, alpha = 0.4, interpolate = TRUE) +
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_fixed()
```



Susan Holmes & Wolfgang Huber: Modern Statistics for Modern Biology

Base R statistical output

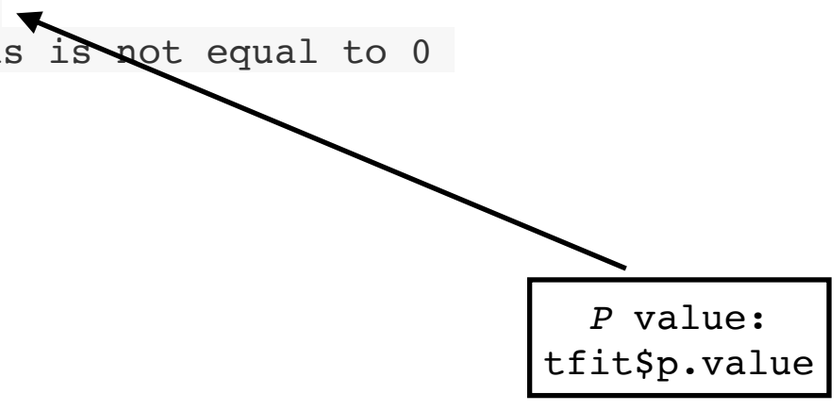
```
tfit <- t.test(1:10, 10:20)
tfit
```

```
##  Welch Two Sample t-test
```

```
## data:  1:10 and 10:20
## t = -6.862, df = 18.998, p-value = 0.000001513
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -12.397677  -6.602323
## sample estimates:
## mean of x mean of y
##       5.5      15.0
```

Base R statistical output

```
tfit <- t.test(1:10, 10:20)
tfit

##  Welch Two Sample t-test

## data:  1:10 and 10:20
## t = -6.862, df = 18.998, p-value = 0.000001513
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -12.397677  -6.602323
## sample estimates:
## mean of x mean of y
##       5.5      15.0
```



Base R statistical output

```
lmfit <- lm(mpg ~ wt, mtcars)
lmfit
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Coefficients:
## (Intercept)           wt
##      37.285       -5.344
```

```
summary(lmfit)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

Base R statistical output

```
lmfit <- lm(mpg ~ wt, mtcars)
lmfit
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Coefficients:
## (Intercept)           wt
##      37.285       -5.344
```

```
summary(lmfit)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

P value:
`summary(lmfit)$coefficients[2, 4]`

New packages are tidy


```
library(broom)

tidy(tfit)
# A tibble: 1 x 10
  estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high method
     <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
1     -9.5       5.5        15     -6.86 1.51e-6      19.0    -12.4     -6.60 Welch…
# … with 1 more variable: alternative <chr>

tidy(lmfit)
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    37.3      1.88      19.9  8.24e-19
2 wt             -5.34     0.559     -9.56 1.29e-10
```

P value:
tfit$p.value
lmfit$p.value[2]

David Robinson & Alex Hayes

```
install.packages("broom")
```

New packages are tidy

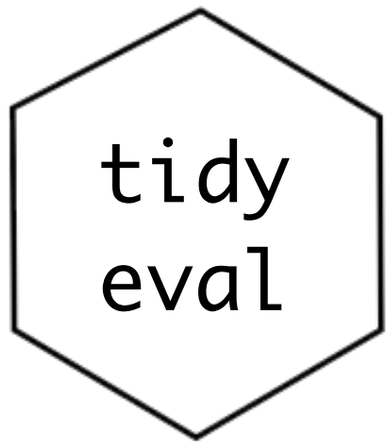

```
glance(lmfit)
## Summary statistics
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic  p.value    df logLik   AIC   BIC
*     <dbl>         <dbl> <dbl>     <dbl>    <dbl> <int>  <dbl> <dbl> <dbl>
1     0.753         0.745  3.05      91.4 1.29e-10     2  -80.0  166.  170.
# ... with 2 more variables: deviance <dbl>, df.residual <int>

augment(lmfit)
## Fitted values and residuals (information about the model)
# A tibble: 32 x 10
   .rownames    mpg    wt .fitted .se.fit .resid   .hat .sigma .cooksd
 * <chr>     <dbl> <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl>   <dbl>
 1 Mazda RX4  21    2.62    23.3   0.634 -2.28  0.0433   3.07 1.33e-2
 2 Mazda RX~  21    2.88    21.9   0.571 -0.920 0.0352   3.09 1.72e-3
 3 Datsun 7~  22.8  2.32    24.9   0.736 -2.09  0.0584   3.07 1.54e-2
 4 Hornet 4~  21.4  3.22    20.1   0.538  1.30  0.0313   3.09 3.02e-3
 5 Hornet S~  18.7  3.44    18.9   0.553 -0.200 0.0329   3.10 7.60e-5
 6 Valiant    18.1  3.46    18.8   0.555 -0.693 0.0332   3.10 9.21e-4
 7 Duster 3~  14.3  3.57    18.2   0.573 -3.91  0.0354   3.01 3.13e-2
 8 Merc 240D  24.4  3.19    20.2   0.539  4.16  0.0313   3.00 3.11e-2
 9 Merc 230   22.8  3.15    20.5   0.540  2.35  0.0314   3.07 9.96e-3
10 Merc 280   19.2  3.44    18.9   0.553  0.300 0.0329   3.10 1.71e-4
# ... with 22 more rows, and 1 more variable: .std.resid <dbl>
```

David Robinson & Alex Hayes

```
install.packages("broom")
```

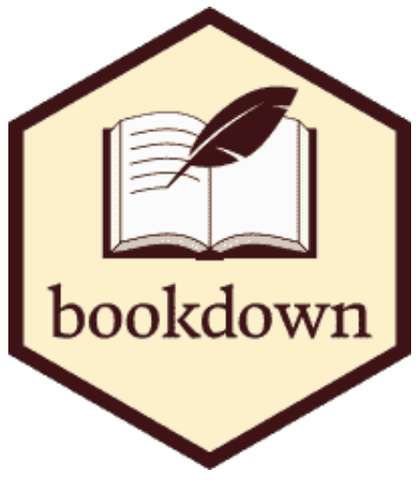
New packages are tidy

Standardizes the interface for fitting models as well as the return values



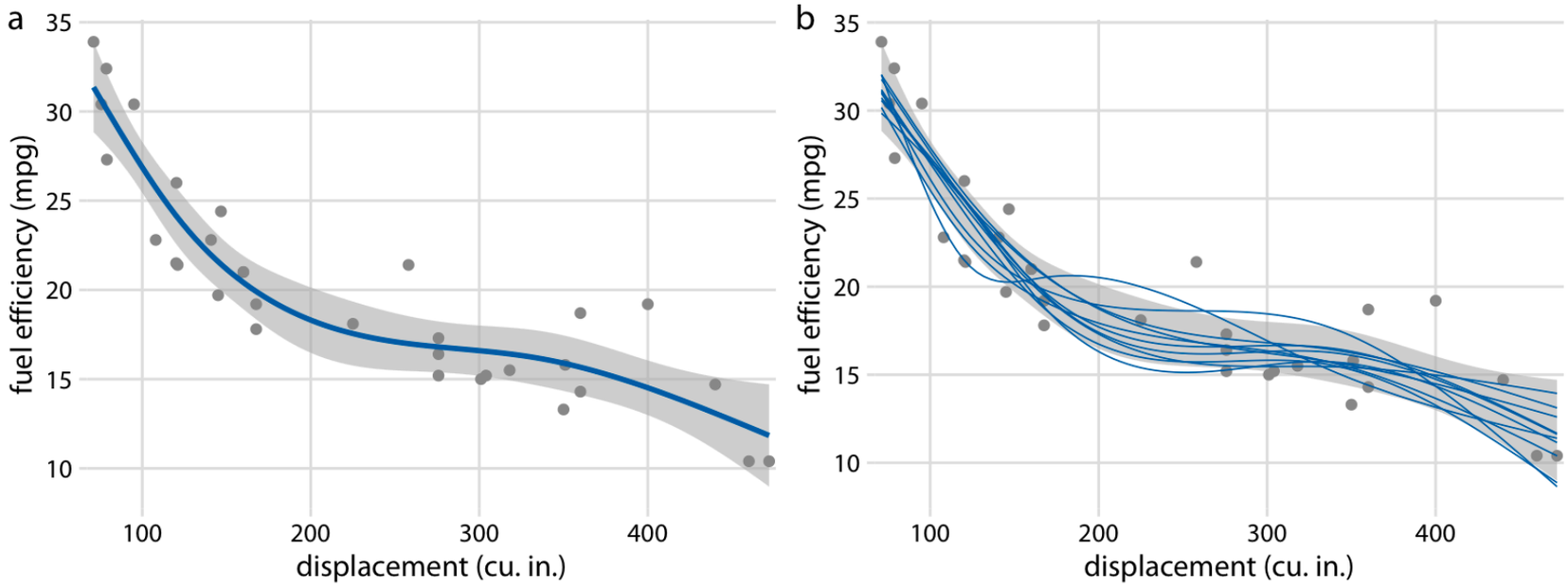
For writing functions around tidyverse pipelines and grammar

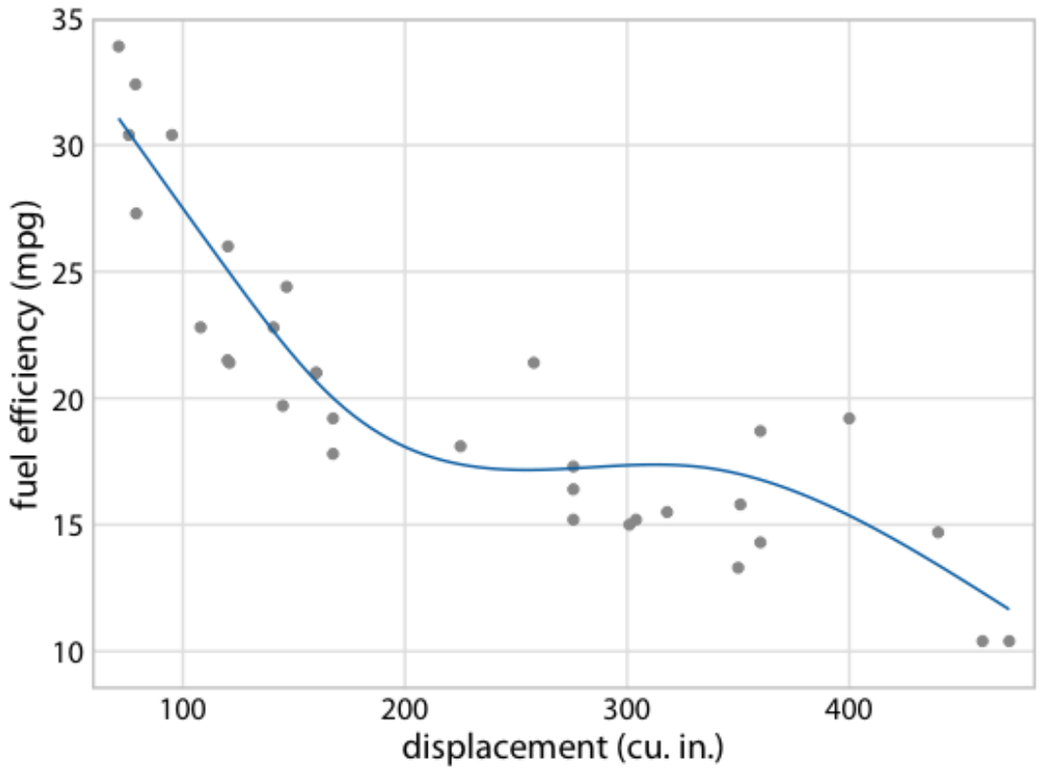
`https://tidyeval.tidyverse.org/`



(not tidy *per se*) Consistent documentation format

Hypothetical outcome plots






```
install.packages("gganimate")
```

Claus Wilke: https://serialmentor.com/dataviz/visualizing-uncertainty.html#hypothetical-outcome-plots

Corporate R takeover

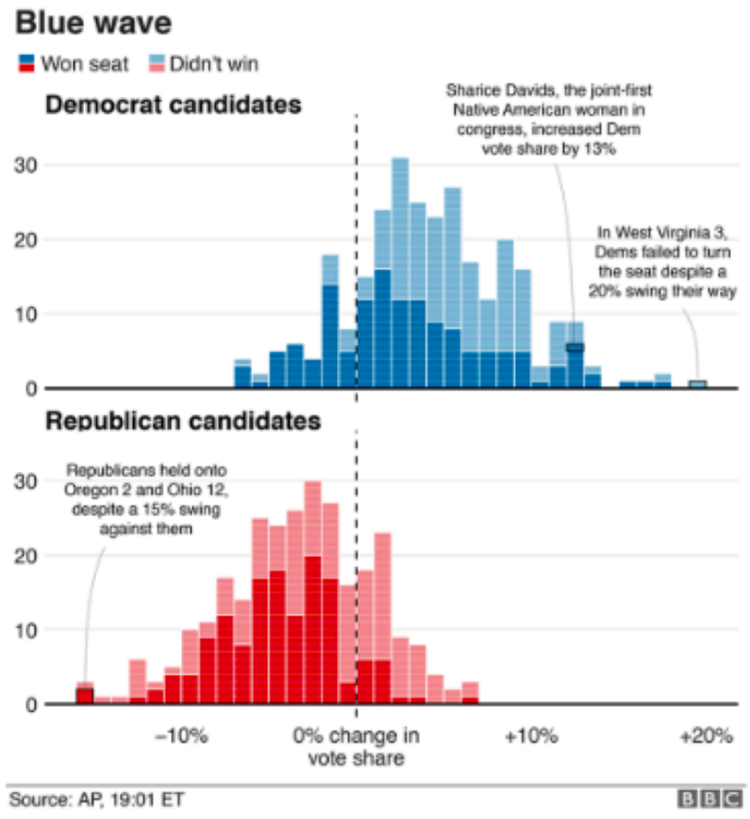
BBC Visual Journalism data team's cookbook for R graphics

Last updated: 2019-01-16

How to create BBC style graphics

At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

The cookbook below should hopefully help anyone who wants to make graphics like these:



We'll get to how you can put together the various elements of these graphics, but **let's get the admin out of the way first...**

```
install.packages("bbplot")
```

Top Tier Companies using R

The following is a list of top brands or large organizations using R.

1. Facebook – For behavior analysis related to status updates and profile pictures.
2. Google – For advertising effectiveness and economic forecasting.
3. Twitter – For data visualization and semantic clustering
4. Microsoft – Acquired Revolution R company and use it for a variety of purposes.
5. Uber – For statistical analysis
6. Airbnb – Scale data science.
7. IBM – Joined R Consortium Group
8. ANZ – For credit risk modeling
9. HP
10. Ford
11. Novartis
12. Roche
13. New York Times – For data visualization
14. Mckinsey
15. BCG
16. Bain

Analytics and Consulting Companies using R

The below list comprises of niche analytics companies as well as consulting companies providing analytics or market research services.

1. A.T. Kearney
2. AbsolutData
3. AC Nielsen
4. Accenture
5. Bain & Company
6. Booz Allen Hamilton
7. Capgemini
8. Convergytics
9. Deloitte Consulting
10. Evalueserve
11. EXL
12. EY
13. Fractal Analytics
14. Gartner
15. Genpact
16. IBM
17. KPMG
18. Latent View
19. Manthan Systems
20. McKinsey & Company
21. Mu Sigma
22. PricewaterhouseCoopers
23. SIBIA Analytics
24. Simplify360
25. SmartCube
26. Target
27. The Boston Consulting Group
28. Tiger Analytics
29. Tower Watson
30. WNS
31. ZS Associate

https://www.r-bloggers.com/companies-using-r/

IT Companies using R

It includes major companies providing IT and professional services using R in India and other parts of the world.

1. Accenture
2. Amadeus IT Group
3. Capgemini
4. Cognizant
5. CSC
6. HCL Technologies
7. Hexaware Technologies
8. HP
9. IBM
10. IGATE
11. Infosys
12. Larsen & Toubro Infotech
13. Microsoft
14. Mindtree
15. Mphasis
16. NIIT Tech
17. Oracle Financial Services Software
18. Paytm
19. Snapdeal
20. R Systems Ltd
21. Tata Consultancy Services
22. Tech Mahindra
23. Wipro

Financial Institutions

It includes major US and European Banks, Insurance Companies and Other financial institutions using R.

1. American Express
2. ANZ
3. Bank of America
4. Barclays Bank
5. Bazaj allianz Insurance
6. Bharti Axa insurance
7. Blackrock
8. Citibank
9. Dun &Bradstreet
10. Fidelity
11. HSBC
12. JP Morgan
13. KeyBank
14. Lloyds Banking
15. RBS
16. Standard Chartered
17. UBS
18. Wells Fargo
19. Goldman Sachs
20. Morgan Stanley
21. PNC Bank
22. Citizens Bank
23. Fifth Third Bank

and

Vicki Boykis
@vboykis

Have been extremely curious about this for a while now, so I decided to create a poll.
"As someone titled 'data scientist' in 2019, I spend most of (60%+) my time:"
("Other") also welcome, add it in the replies.

♡ 195 10:17 AM - Jan 28, 2019

6% Picking features/models

67% Cleaning data/Moving data

4% Deploying models in prod

23% Analyzing/presenting data

2,116 votes • Final results

and

Sam Langton
@sh_langton

Follow

> It's only recently become clear to me that novices can help older and more experienced #rstats users simply by (1) knowing #tidyverse fundamentals, and (2) knowing how to Google them. The beginner/expert sharing of skills can be a two-way street if the latter is open to it.

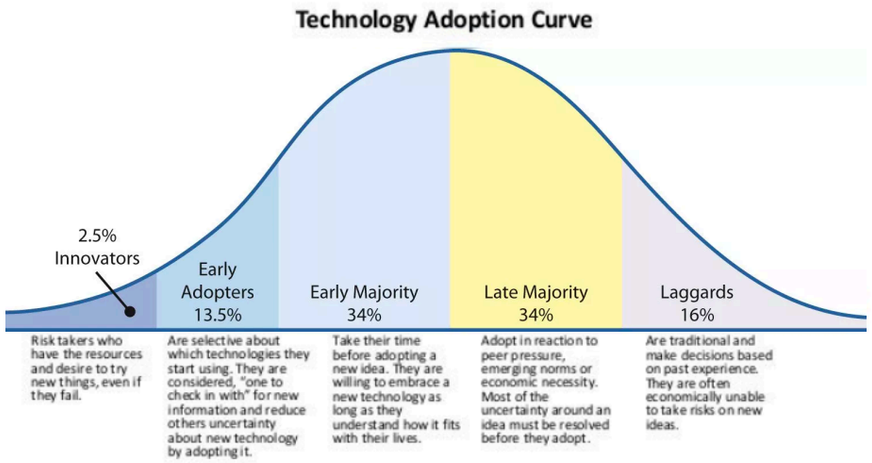
6:33 AM - 4 Feb 2019

Claus Wilke
@ClausWilke

Following

Replying to @hadleywickham @thosjleeper

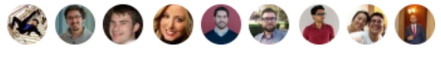
I think this graphic applies here. For tidyverse adoption we're maybe at the beginning of "late majority" at this time. We need to keep in mind that this is still very new technology, even if it seems so familiar to us. dplyr is ~5 years old. tidyverse less than 3.



Popularized by Everett Rogers in the book *Diffusion of Innovations*, people tend to adopt new technologies at varying rates. Their relative speed of adoption follows a bell curve, with the primary difference being individuals' psychological disposition to new ideas.

11:18 AM - 5 Feb 2019

4 Retweets **25** Likes



2 4 25

Tweet your reply

Hadley Wickham @hadleywickham · Feb 5

Replying to @ClausWilke @thosjleeper

I think we're probably still in early majority

8

Practicality (finding help!)

The internet will make those bad words go away

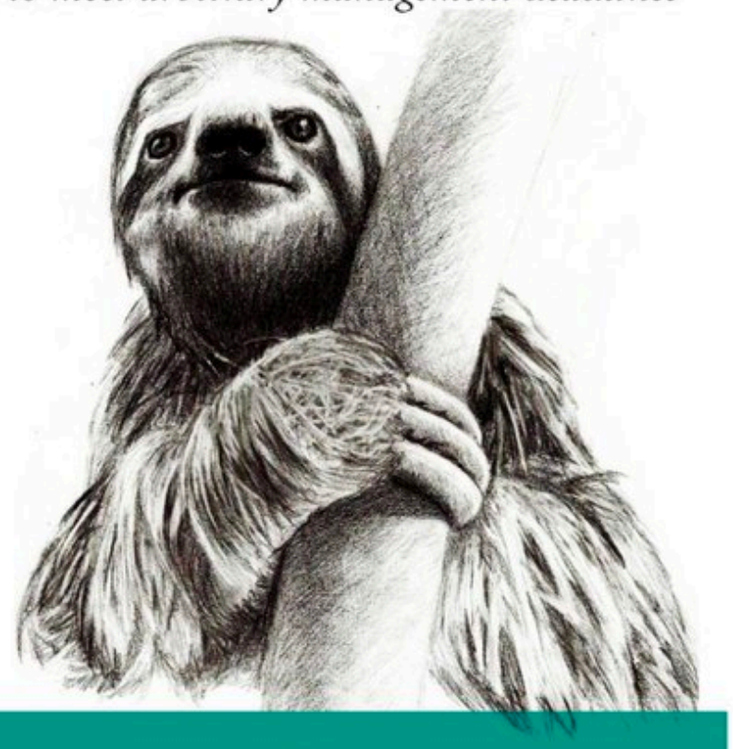
Essential

Googling the Error Message

O RLY?

The Practical Developer
@ThePracticalDev

Cutting corners to meet arbitrary management deadlines



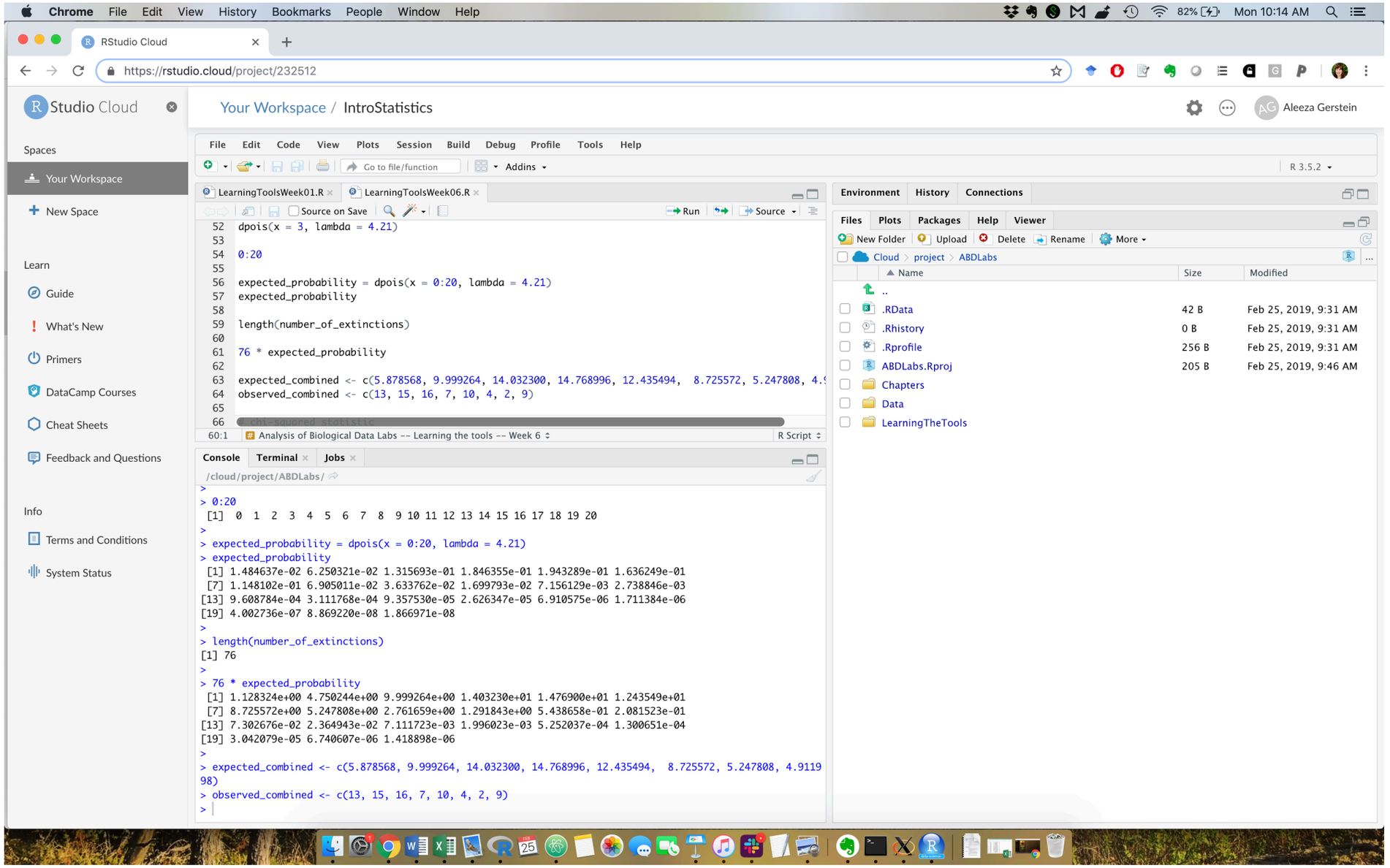
Essential

Copying and Pasting from Stack Overflow

O'REILLY®

The Practical Developer
@ThePracticalDev

RStudio Cloud



Mel Gregory: https://resources.rstudio.com/rstudio-conf-2018/learning-r-with-rstudio-cloud-mel-gregory

RStudio Cloud

DATA WITH SOME SCIENCE

Feb 1 Modern Master

Aesthetics

SCIENCE DOES NOT HAVE TO BE A SHAMBLING, UGLY THING. JJ ALLAIRE AND YIHUI XIE HAVE CREATED A LOVELY TEMPLATE BASED ON THE WORK OF EDWARD TUFTE. IN THIS POST, WE EXPLORE THE INS AND OUTS OF FOLLOWING IN A MASTER'S FOOTSTEPS.

> “ The commonality between science and art is in trying to see profoundly – to develop strategies of seeing and showing. ”
>
> — EDWARD TUFTE

Edward Rolf Tufte is an American statistician and professor emeritus of political science, statistics, and computer science at Yale University. He is noted for his writings on information design and as a pioneer in the field of data visualization. If you haven't looked at any of his books,

Mel Gregory: https://resources.rstudio.com/rstudio-conf-2018/learning-r-with-rstudio-cloud-mel-gregory

Now that that's over with, we can tell you about the R package "tufte". It makes it incredibly easy to create beautiful RMarkdown documents. Here's an example project that we put together that demonstrates the basic features. The project is hosted on RStudio Cloud - you'll need an account to open it - but don't let that deter you. An account is free, and you'll find that it's an amazing resource for doing, sharing, teaching and learning data science.

> “ The leading edge in evidence presentation is in science; the leading edge in beauty is in high art. ”
>
> — EDWARD TUFTE

Mel Gregory: https://resources.rstudio.com/rstudio-conf-2018/learning-r-with-rstudio-cloud-mel-gregory

https://rstudio.cloud/project/9882

RStudio Cloud

Spaces

Your Workspace

New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

Your Workspace / Tufte Example Robby Shaver

TEMPORARY

Save a Permanent Copy

Aleeza Gerstein

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | R 3.5.2

Tufte1.Rmd

Knit | Insert | Run

```
17
18 ```{r setup, include=FALSE}
19 library(tufte)
20 # invalidate cache when the tufte version changes
21 knitr::opts_chunk$set(tidy = FALSE, cache.extra = packageVersion('tufte'))
22 options(htmltools.dir.version = FALSE)
23 ```
24
25 # Introduction
26
27 The Tufte handout style is a style that Edward Tufte uses in his books and handouts.
Tufte's style is known for its extensive use of sidenotes, tight integration of graphics
with text, and well-set typography. This style has been implemented in LaTeX and
HTML/CSS^[See Github repositories
[tufte-latex](https://github.com/tufte-latex/tufte-latex) and
[tufte-css](https://github.com/edwardtufte/tufte-css)], respectively. We have ported both
implementations into the [**tufte** package](https://github.com/rstudio/tufte). If you
want LaTeX/PDF output, you may use the `tufte_handout` format for handouts, and
`tufte_book` for books. For HTML output, use `tufte_html`. These formats can be either
specified in the YAML metadata at the beginning of an R Markdown document (see an example
```

13:26 Tufte Example R Markdown

Console Terminal R Markdown Jobs

/cloud/project/

```
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

Environment History Connections

Import Dataset | List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder | Upload | Delete | Rename | More

Cloud > project

Name	Size	Modified
..		
skeleton.bib	603 B	Apr 5, 2018, 1:53 PM
Tufte1.html	1.1 MB	Apr 5, 2018, 1:53 PM
Tufte1.Rmd	12.5 KB	Apr 16, 2018, 1:29 PM
Tufte1_cache		
Tufte1_files		

Mel Gregory: https://resources.rstudio.com/rstudio-conf-2018/learning-r-with-rstudio-cloud-mel-gregory

Tufte Example

An implementation in R Markdown

JJ Allaire and Yihui Xie

2018-04-05

Introduction

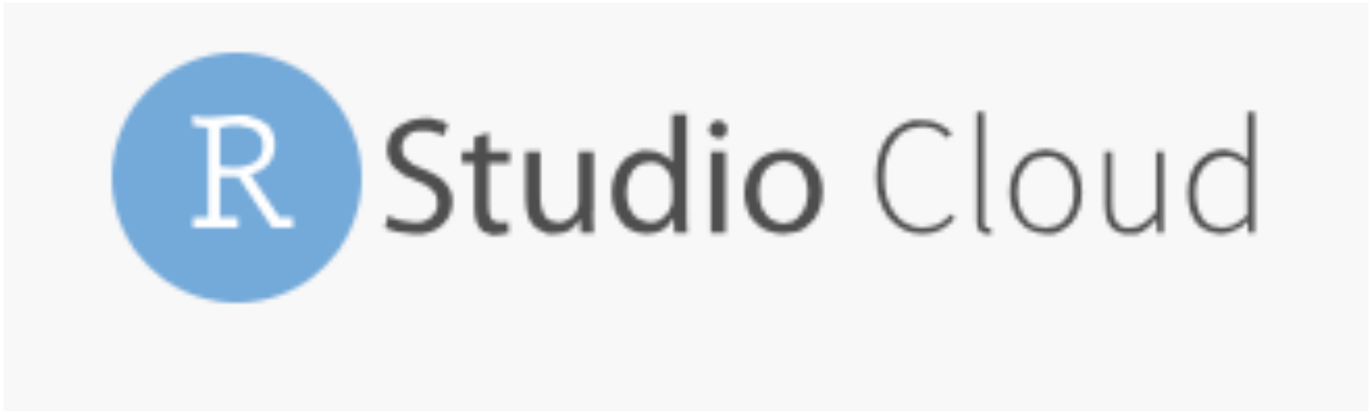
The Tufte handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, tight integration of graphics with text, and well-set typography. This style has been implemented in LaTeX and HTML/CSS[1], respectively. We have ported both implementations into the **tufte** package. If you want LaTeX/PDF output, you may use the `tufte_handout` format for handouts, and `tufte_book` for books. For HTML output, use `tufte_html`. These formats can be either specified in the YAML metadata at the beginning of an R Markdown document (see an example below), or passed to the `rmarkdown::render()` function. See Allaire et al. (2017) more information about **rmarkdown**.

[1] See Github repositories tufte-latex and tufte-css

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, and Winston Chang. 2017. *Rmarkdown: Dynamic Documents for R*. https://CRAN.R-project.org/package=rmarkdown.

```
---
title: "An Example Using the Tufte Style"
author: "John Smith"
output:
  tufte::tufte_handout: default
  tufte::tufte_html: default
---
```

There are two goals of this package:



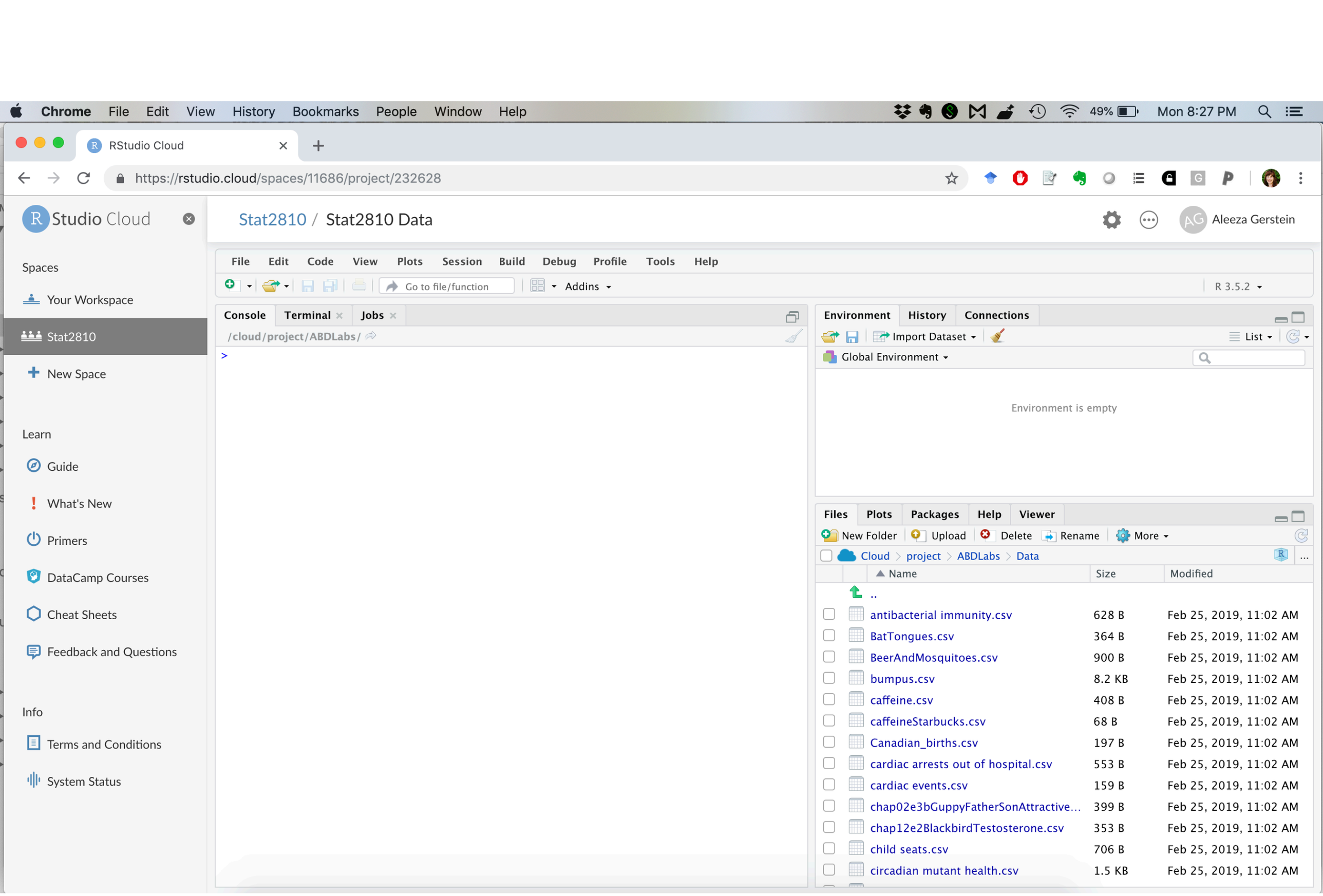
In a classroom:

No time spent on setup & installation

Easily distribute scripts, files, assignments

Access students work easily (scripts, RMD)

Mel Gregory: https://resources.rstudio.com/rstudio-conf-2018/learning-r-with-rstudio-cloud-mel-gregory


Chrome File Edit View History Bookmarks People Window Help
49% Mon 8:27 PM
RStudio Cloud
https://rstudio.cloud/spaces/11686/project/232628
Studio Cloud
Stat2810 / Stat2810 Data
Aleeza Gerstein
Spaces
Your Workspace
Stat2810
New Space
Learn
Guide
What's New
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 3.5.2
Console
Terminal
Jobs
/cloud/project/ABDLabs/
>
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
ABDLabs
Data
Name
Size
Modified
..
antibacterial immunity.csv 628 B Feb 25, 2019, 11:02 AM
BatTongues.csv 364 B Feb 25, 2019, 11:02 AM
BeerAndMosquitoes.csv 900 B Feb 25, 2019, 11:02 AM
bumpus.csv 8.2 KB Feb 25, 2019, 11:02 AM
caffeine.csv 408 B Feb 25, 2019, 11:02 AM
caffeineStarbucks.csv 68 B Feb 25, 2019, 11:02 AM
Canadian_births.csv 197 B Feb 25, 2019, 11:02 AM
cardiac arrests out of hospital.csv 553 B Feb 25, 2019, 11:02 AM
cardiac events.csv 159 B Feb 25, 2019, 11:02 AM
chap02e3bGuppyFatherSonAttractive... 399 B Feb 25, 2019, 11:02 AM
chap12e2BlackbirdTestosterone.csv 353 B Feb 25, 2019, 11:02 AM
child seats.csv 706 B Feb 25, 2019, 11:02 AM
circadian mutant health.csv 1.5 KB Feb 25, 2019, 11:02 AM

Chrome File Edit View History Bookmarks People Window Help 100% Thu 11:24 AM

RStudio Cloud

https://rstudio.cloud/spaces/11686/project/232628

Studio Cloud

Stat2810 / Stat2810 Data

Aleeza Gerstein

Spaces

Your Workspace

Stat2810

New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | R 3.5.2

Console Terminal Jobs

/cloud/project/ABDLabs/

>

Environment History Connections

Import Dataset | List

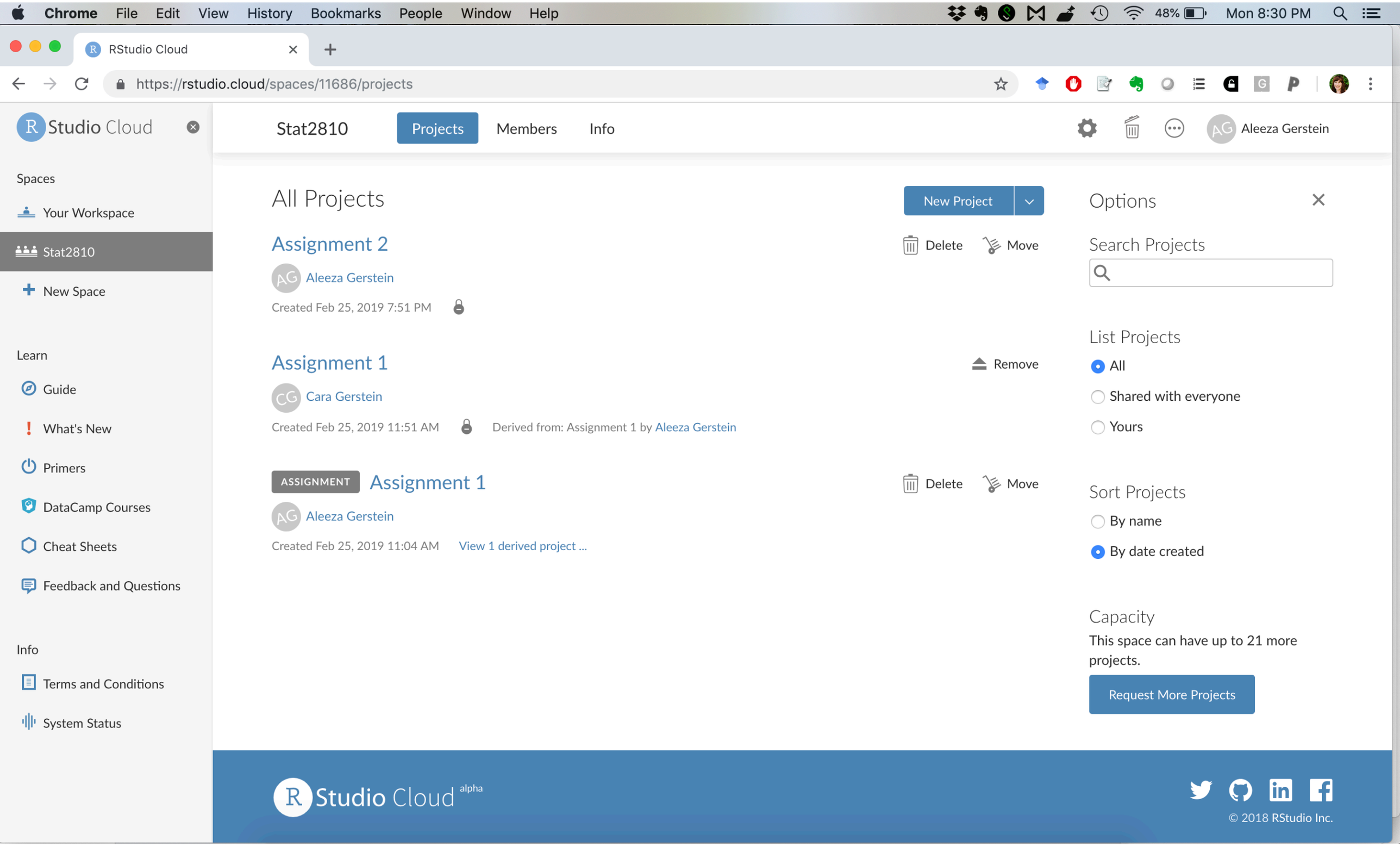
Global Environment

Environment is empty

Files Plots Packages Help Viewer

Install | Update | Packrat

	Name	Description	Version
System Library			
☑	base	The R Base Package	3.5.2
☐	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-20
☐	class	Functions for Classification	7.3-14
☐	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.7-1
☐	codetools	Code Analysis Tools for R	0.2-15
☑	compiler	The R Compiler Package	3.5.2
☑	datasets	The R Datasets Package	3.5.2
☐	foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...	0.8-71
☑	graphics	The R Graphics Package	3.5.2
☑	grDevices	The R Graphics Devices and Support for Colours and Fonts	3.5.2
☐	grid	The Grid Graphics Package	3.5.2
☐	KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-15
☐	lattice	Trellis Graphics for R	0.20-38
☐	MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-51.1
☐	Matrix	Sparse and Dense Matrix Classes and Methods	1.2-15
☑	methods	Formal Methods and Classes	3.5.2
☐	mgcv	Mixed GAM Computation Vehicle with Automatic Smoothness Estimation	1.8-26
☐	nlme	Linear and Nonlinear Mixed Effects Models	3.1-137
☐	nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models	7.3-12
☐	parallel	Support for Parallel computation in R	3.5.2
☐	rpart	Recursive Partitioning and Regression Trees	4.1-13
☐	spatial	Functions for Kriging and Point Pattern Analysis	7.3-11
☐	splines	Regression Spline Functions and Classes	3.5.2
☑	stats	The R Stats Package	3.5.2
☐	stats4	Statistical Functions using S4 Classes	3.5.2


Chrome File Edit View History Bookmarks People Window Help
48% Mon 8:30 PM
RStudio Cloud
https://rstudio.cloud/spaces/11686/projects
RStudio Cloud
Spaces
Your Workspace
Stat2810
New Space
Learn
Guide
What's New
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
Stat2810
Projects
Members
Info
Aleeza Gerstein
All Projects
New Project
Assignment 2
Aleeza Gerstein
Created Feb 25, 2019 7:51 PM
Delete
Move
Assignment 1
Cara Gerstein
Created Feb 25, 2019 11:51 AM
Derived from: Assignment 1 by Aleeza Gerstein
Remove
ASSIGNMENT
Assignment 1
Aleeza Gerstein
Created Feb 25, 2019 11:04 AM
View 1 derived project ...
Delete
Move
Options
Search Projects
List Projects
All
Shared with everyone
Yours
Sort Projects
By name
By date created
Capacity
This space can have up to 21 more projects.
Request More Projects
RStudio Cloud alpha
© 2018 RStudio Inc.

Chrome File Edit View History Bookmarks People Window Help 55% Mon 8:03 PM

RStudio Cloud

https://rstudio.cloud/spaces/11686/project/232635

RStudio Cloud

Spaces

Your Workspace

Stat2810

New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

Stat2810 / Assignment 1

Aleeza Gerstein

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | R 3.5.2

Assignment 1.Rmd

Knit | Insert | Run

```
1  ---
2  title: "Assignment 1"
3  output: html_document
4  ---
5
6  ```{r setup, include=FALSE}
7  knitr::opts_chunk$set(echo = TRUE)
8  ```
9
10 ## STAT 2810 - Assignment #1
11
12 #### Introduction
13
14 Researchers have collected data on world record times in the 100 meter freestyle for both men
   and women since 1905. Our task is to analyze that data.
15
16 ```{r, message=FALSE}
```

17:1 | Chunk 2 | R Markdown

Console Terminal R Markdown Jobs

/cloud/project/

>

Environment History Connections

Import Dataset | List

Global Environment

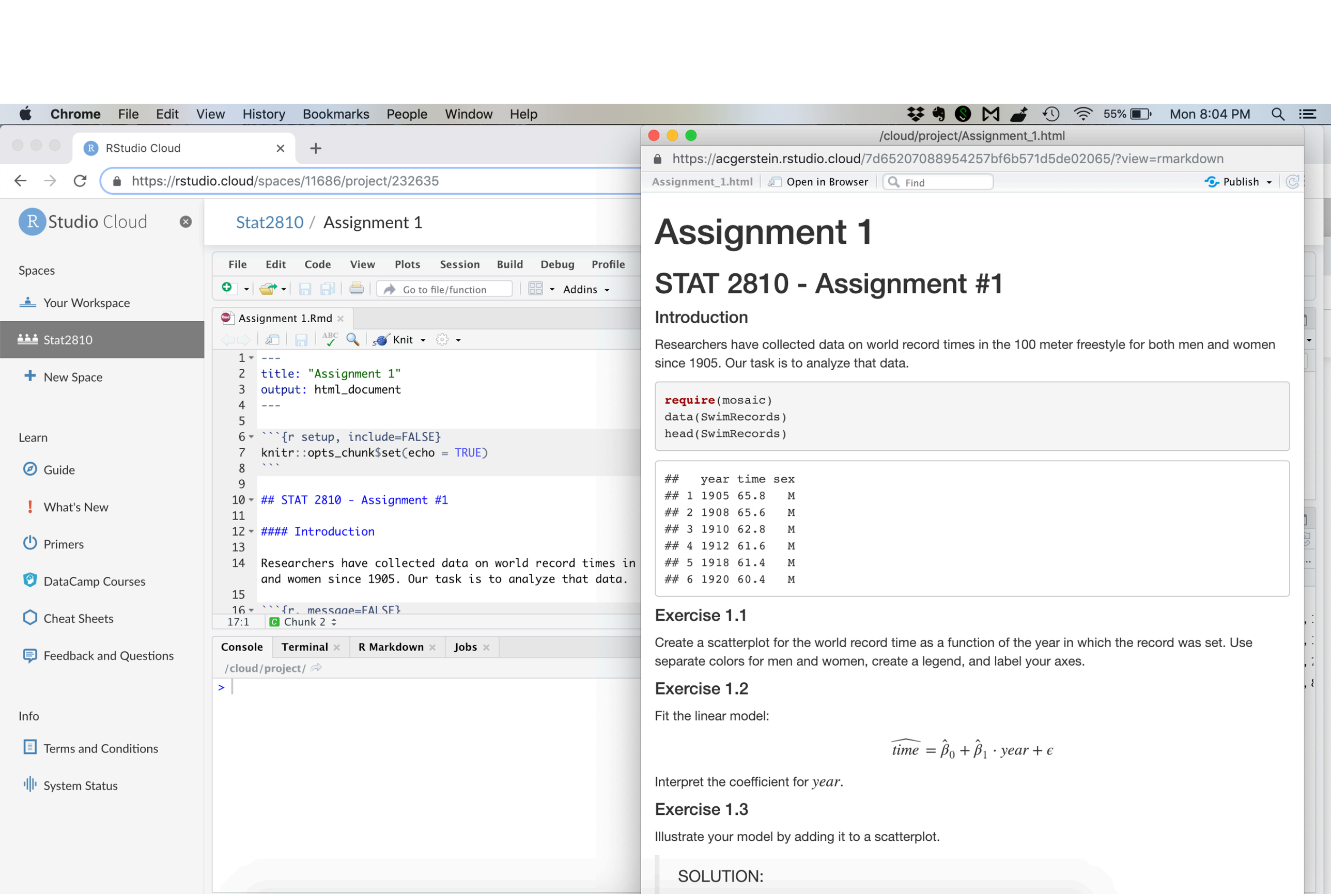
Environment is empty

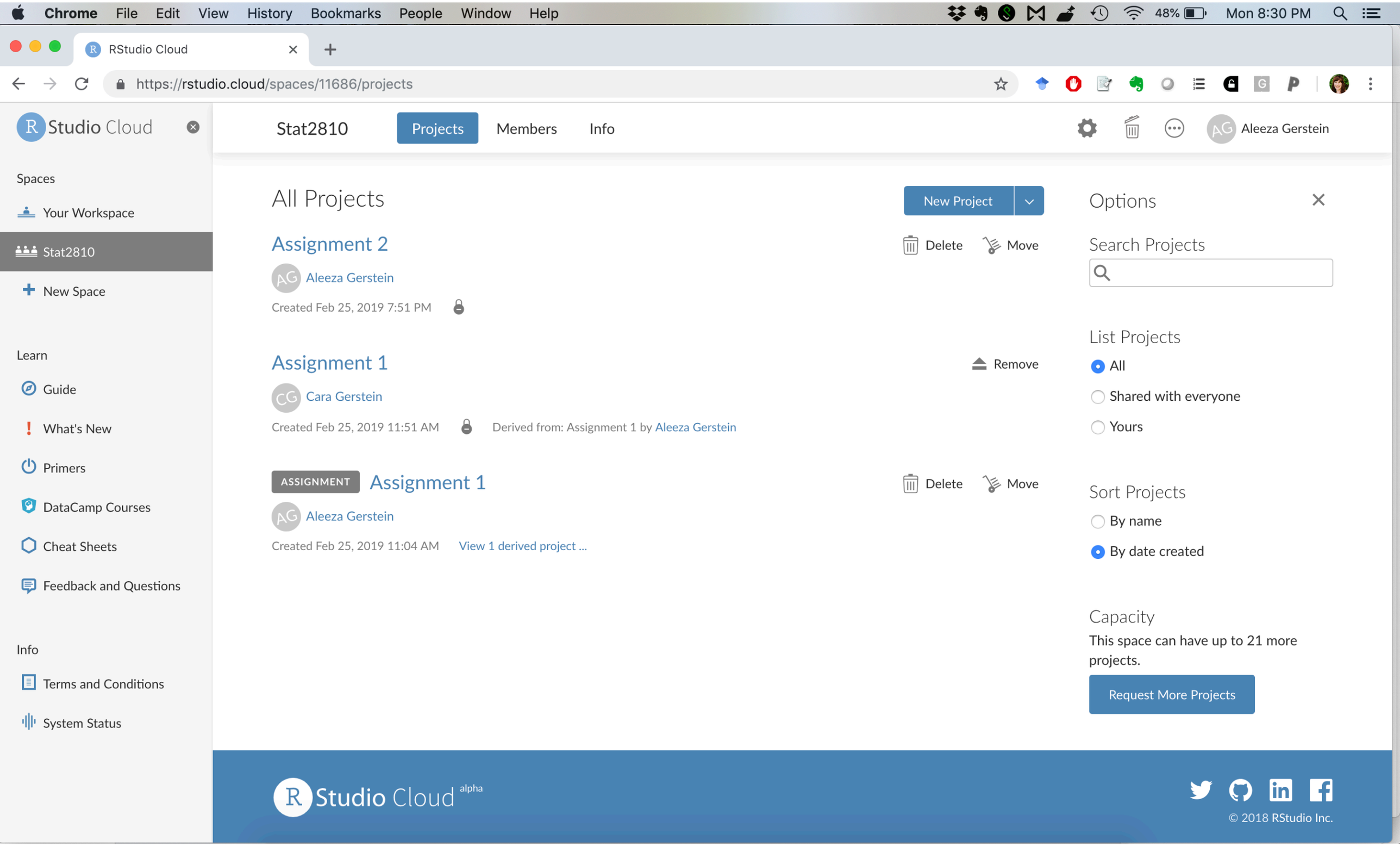
Files Plots Packages Help Viewer

New Folder | Upload | Delete | Rename | More

Cloud > project

Name	Size	Modified
..		
.Rhistory	0 B	Feb 25, 2019,
Assignment 1.Rmd	1.3 KB	Feb 25, 2019,
project.Rproj	205 B	Feb 25, 2019,
Assignment_1.html	616.3 KB	Feb 25, 2019,


Chrome File Edit View History Bookmarks People Window Help
55% Mon 8:04 PM
RStudio Cloud
https://rstudio.cloud/spaces/11686/project/232635
Studio Cloud
Spaces
Your Workspace
Stat2810
New Space
Learn
Guide
What's New
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
Stat2810 / Assignment 1
File Edit Code View Plots Session Build Debug Profile
Go to file/function
Addins
Assignment 1.Rmd
Knit
1 ---
2 title: "Assignment 1"
3 output: html_document
4 ---
5
6 ```{r setup, include=FALSE}
7 knitr::opts_chunk$set(echo = TRUE)
8 ```
9
10 ## STAT 2810 - Assignment #1
11
12 #### Introduction
13
14 Researchers have collected data on world record times in
and women since 1905. Our task is to analyze that data.
15
16 ```{r, message=FALSE}
17:1 Chunk 2
Console Terminal R Markdown Jobs
/cloud/project/
>
/cloud/project/Assignment_1.html
https://acgerstein.rstudio.cloud/7d65207088954257bf6b571d5de02065/?view=rmarkdown
Assignment_1.html Open in Browser Find Publish
Assignment 1
STAT 2810 - Assignment #1
Introduction
Researchers have collected data on world record times in the 100 meter freestyle for both men and women since 1905. Our task is to analyze that data.
require(mosaic)
data(SwimRecords)
head(SwimRecords)
year time sex
1 1905 65.8 M
2 1908 65.6 M
3 1910 62.8 M
4 1912 61.6 M
5 1918 61.4 M
6 1920 60.4 M
Exercise 1.1
Create a scatterplot for the world record time as a function of the year in which the record was set. Use separate colors for men and women, create a legend, and label your axes.
Exercise 1.2
Fit the linear model:
$$\widehat{time} = \hat{\beta}_0 + \hat{\beta}_1 \cdot year + \epsilon$$
Interpret the coefficient for $year$.
Exercise 1.3
Illustrate your model by adding it to a scatterplot.
SOLUTION:


Chrome File Edit View History Bookmarks People Window Help
48% Mon 8:30 PM
RStudio Cloud
https://rstudio.cloud/spaces/11686/projects
RStudio Cloud
Spaces
Your Workspace
Stat2810
New Space
Learn
Guide
What's New
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions
Info
Terms and Conditions
System Status
Stat2810
Projects
Members
Info
Aleeza Gerstein
All Projects
New Project
Assignment 2
Aleeza Gerstein
Created Feb 25, 2019 7:51 PM
Delete
Move
Assignment 1
Cara Gerstein
Created Feb 25, 2019 11:51 AM
Derived from: Assignment 1 by Aleeza Gerstein
Remove
ASSIGNMENT
Assignment 1
Aleeza Gerstein
Created Feb 25, 2019 11:04 AM
View 1 derived project ...
Delete
Move
Options
Search Projects
List Projects
All
Shared with everyone
Yours
Sort Projects
By name
By date created
Capacity
This space can have up to 21 more projects.
Request More Projects
RStudio Cloud alpha
© 2018 RStudio Inc.

Spaces
Your Workspace
Stat2810
New Space

Learn
Guide
What's New
Primers
DataCamp Courses
Cheat Sheets
Feedback and Questions

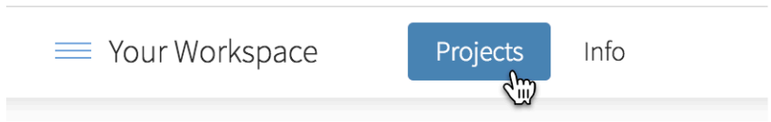
Info
Terms and Conditions
System Status

Studio Cloud Guide

Get started with RStudio Cloud.

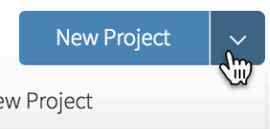
Projects

A project is the fundamental unit of work on RStudio Cloud. It encapsulates your R code, packages and data files and provides isolation from other analyses. If you are familiar with projects in the desktop RStudio IDE, an RStudio Cloud project is the same thing, plus some additional metadata for access and sharing.



To create a new project from scratch, simply press the New Project button from the Projects area. Your new project will open in the RStudio IDE.

To create a new project from an existing git repository, press the down arrow on the right side of the New Project button, and choose 'New Project from Git Repo' from the menu that appears. Note that your git credentials need to be entered each time you create a new project and are only cached for 15 minutes by default. See Working with Git below for more info on working with git.



Spaces

Your Workspace

Stat2810

New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

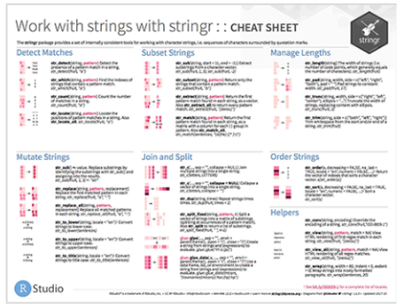
Feedback and Questions

Info

Terms and Conditions

System Status

Work with Strings

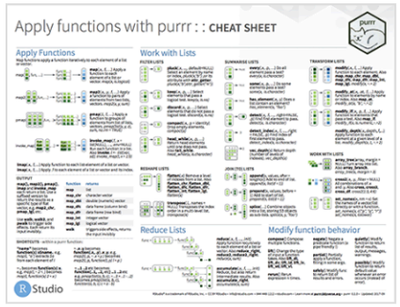


The stringr package provides an easy to use toolkit for working with strings, i.e. character data, in R. This cheatsheet guides you through stringr's functions for manipulating strings. The back page provides a concise reference to regular expresssions, a mini-language for describing, finding, and matching patterns in strings.

Updated October 2017

Download

Apply Functions

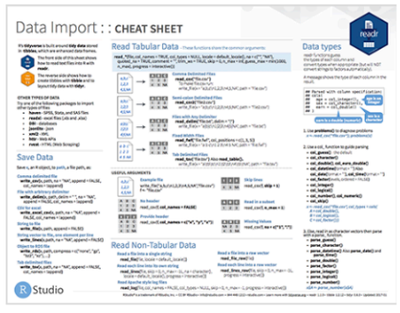


The purrr package makes it easy to work with lists and functions. This cheatsheet will remind you how to manipulate lists with purrr as well as how to apply functions iteratively to each element of a list or vector. The back of the cheatsheet explains how to work with list-columns. With list columns, you can use a simple data frame to organize any collection of objects in R.

Updated September 2017

Download

Data Import

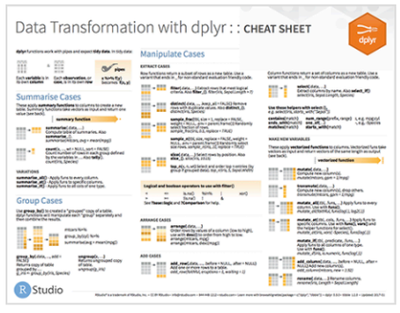


The Data Import cheat sheet reminds you how to read in flat files with http://readr.tidyverse.org/, work with the results as tibbles, and reshape messy data with tidyr. Use tidyr to reshape your tables into tidy data, the data format that works the most seamlessly with R and the tidyverse.

Updated January 2017

Download

Data Transformation

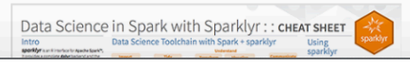


dplyr provides a grammar for manipulating tables in R. This cheat sheet will guide you through the grammar, reminding you how to select, filter, arrange, mutate, summarise, group, and join data frames and tibbles.

Updated January 2017

Download

Sparklyr



Sparklyr provides an R interface to Apache Spark, a fast and general engine for

R Markdown

R Markdown is an authoring format that makes it easy to write reusable reports with

Spaces

Your Workspace

Stat2810

New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

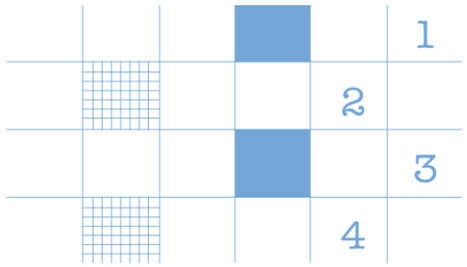
Studio Primers

Learn data science basics with the interactive tutorials below.

The Basics

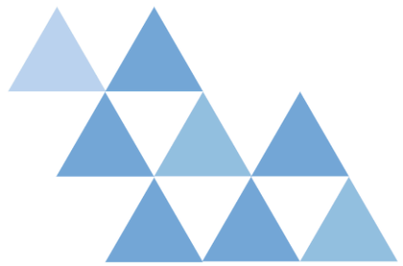
Start here to learn the skills that you will rely on in every analysis (and every primer that follows): how to inspect, visualize, subset, and transform your data, as well as how to run code.

Work with Data



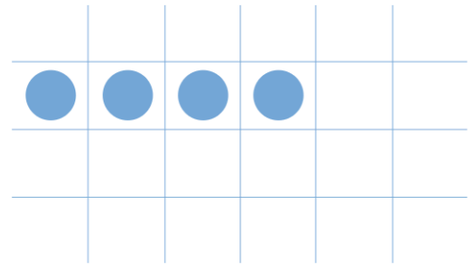
Learn the most important data handling skills in R: how to extract values from a table, subset tables, calculate summary statistics, and derive new variables.

Visualize Data



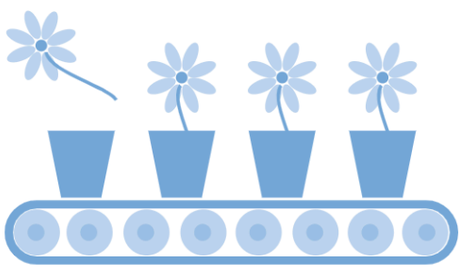
Learn how to use ggplot2 to make any type of plot with your data. Then learn the best ways to visualize patterns within values and relationships between variables.

Tidy Your Data



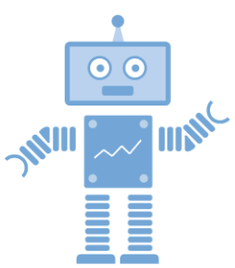
Unlock the tidyverse by learning how to make and use tidy data, the data format designed for R.

Iterate



Master a core programming paradigm with the purrr package: for each ____ do ____.

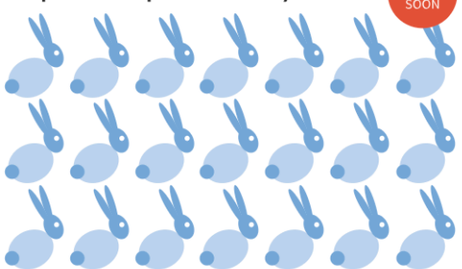
Write Functions



Functions are the key to programming in R. This primer will teach you how to write and use your own reusable functions.

Report Reproducibly

COMING SOON



Learn to report, reproduce, and parameterize your work with the best authoring format for Data Science: R Markdown.

Build Interactive Web Apps

COMING SOON

Say hello to Shiny, R's package for building interactive web apps. Learn to turn your analyses into elegant tools to share with others

DataCamp Courses

DataCamp offers more than 60 courses in R. Click any course below to learn more.

Introduction to R

Master the basics of data analysis by manipulating common data structures such as vectors, matrices and data frames.

4 hours

JONATHAN CORNELISSEN
Co-founder and CEO of DataCamp

Data Analysis in R, the data.table Way

Master core concepts in data manipulation such as subsetting, updating, indexing and joining your data using data.table.

4 hours

MATT DOWLE
Author of data.table

Data Manipulation in R with dplyr

Master techniques for data manipulation using the select, mutate, filter, arrange, and summarise functions in dplyr.

4 hours

GARRETT GROLEMUND
Data Scientist at RStudio

Data Visualization in R with ggvis

Learn to create interactive graphs to display distributions, relationships, model fits, and more using ggvis.

4 hours

GARRETT GROLEMUND
Data Scientist at RStudio

Reporting with R Markdown

Learn to create interactive analyses and automated reports with R Markdown.

3 hours

GARRETT GROLEMUND
Data Scientist at RStudio

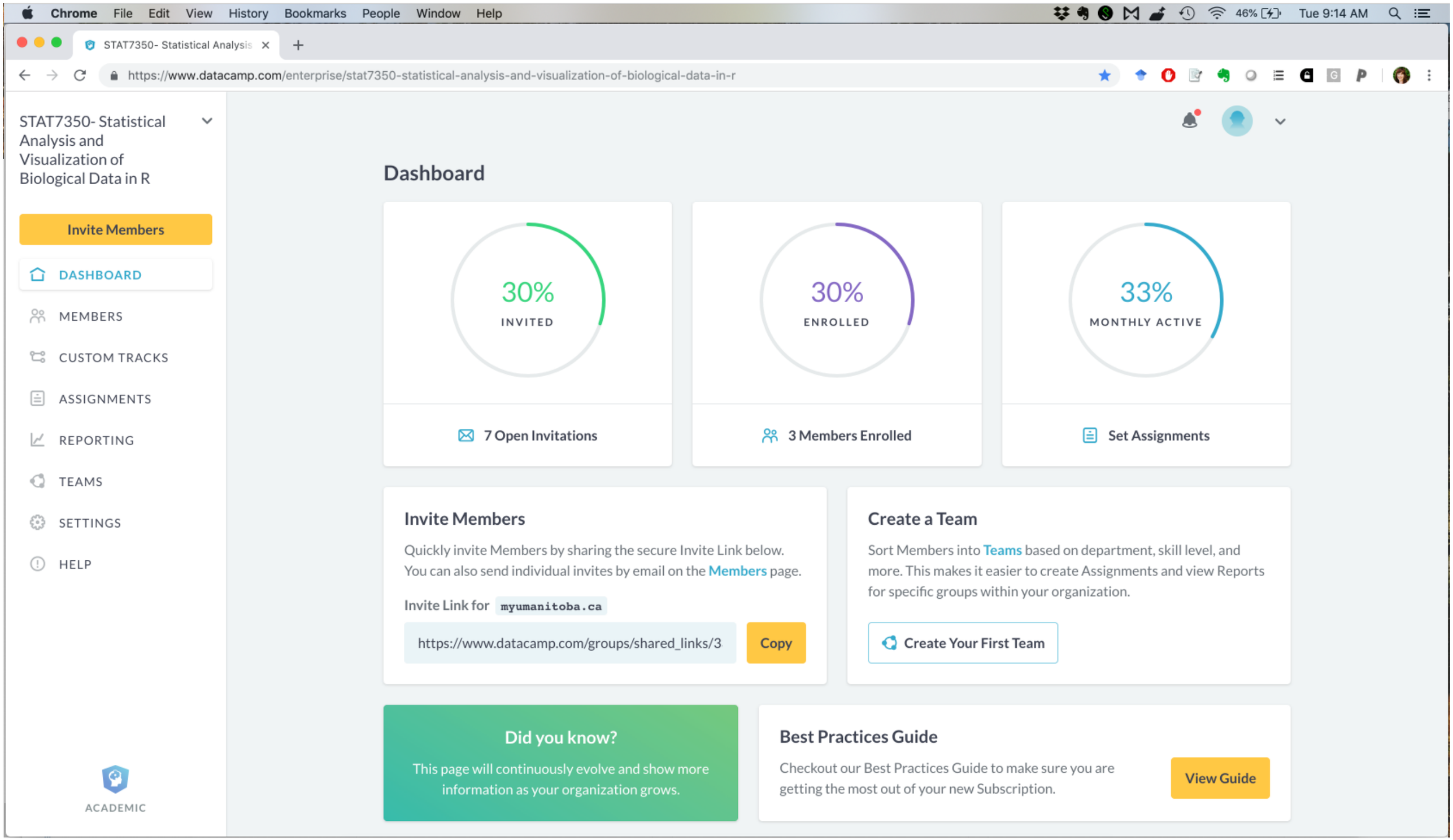
Intermediate R

Continue your journey to become an R ninja by learning about conditional statements, loops, and vector functions.

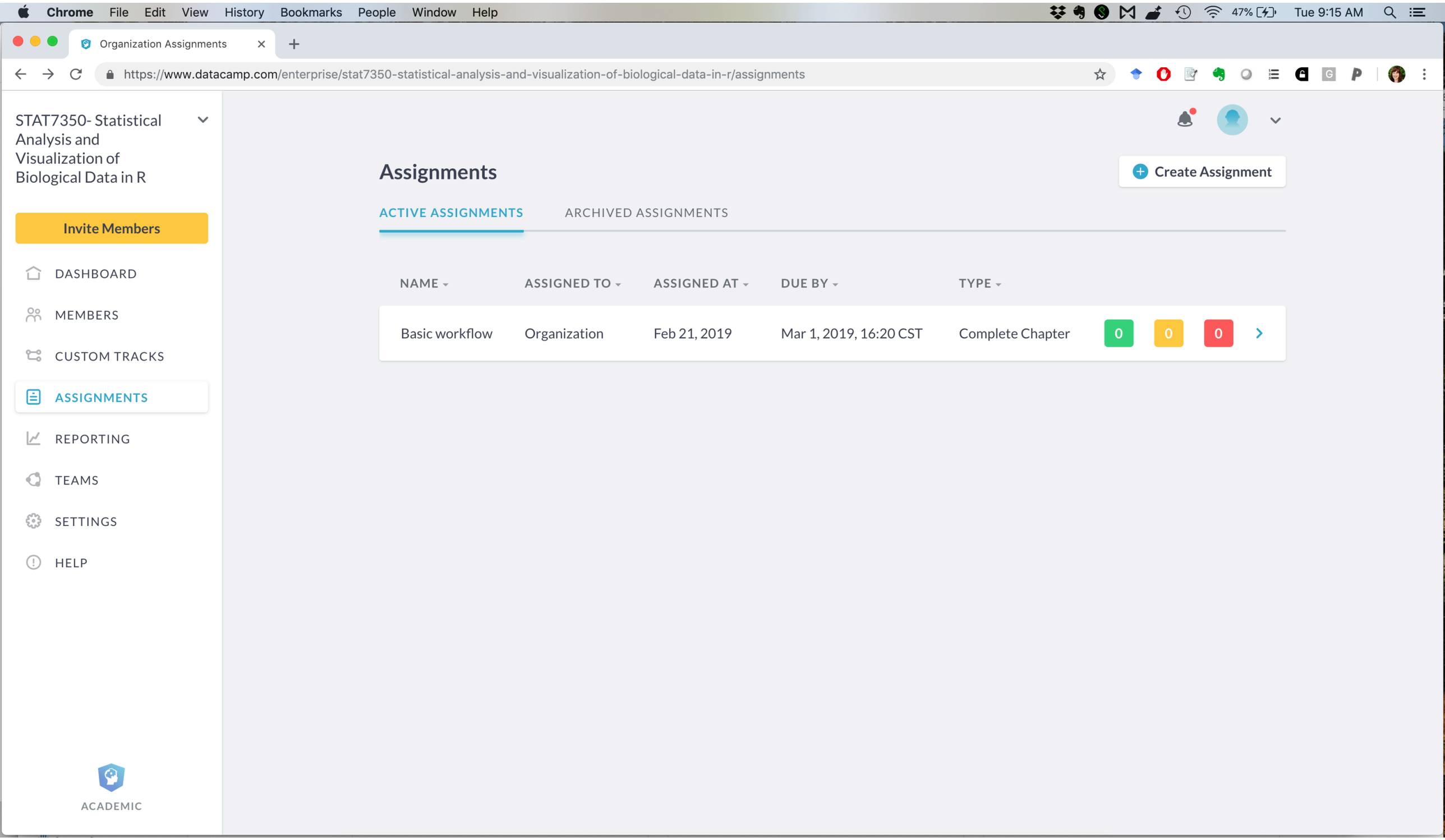
6 hours

FILIP SCHOUWENAARS
Data Science Instructor at DataCamp

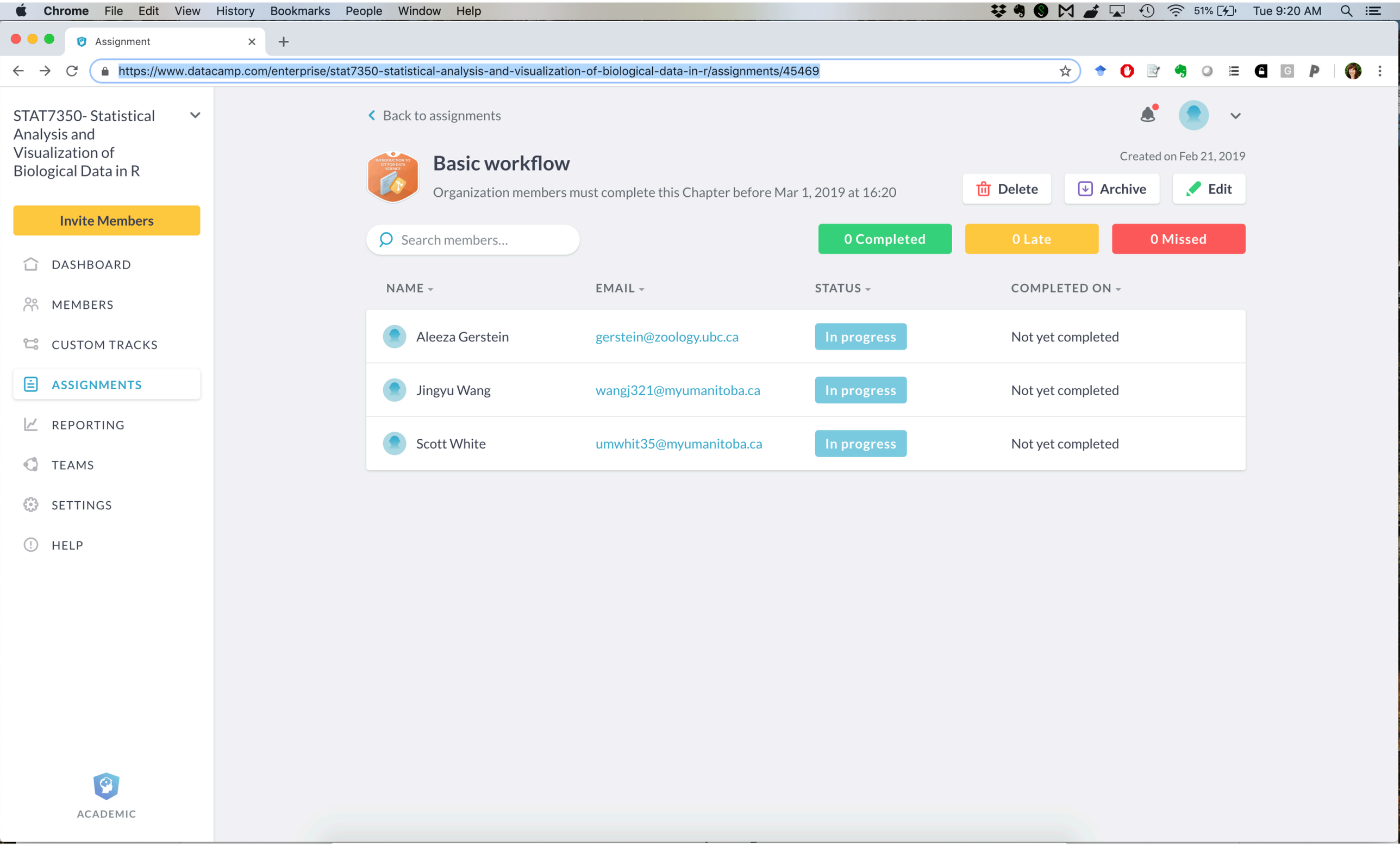
DataCamp for Stat7350



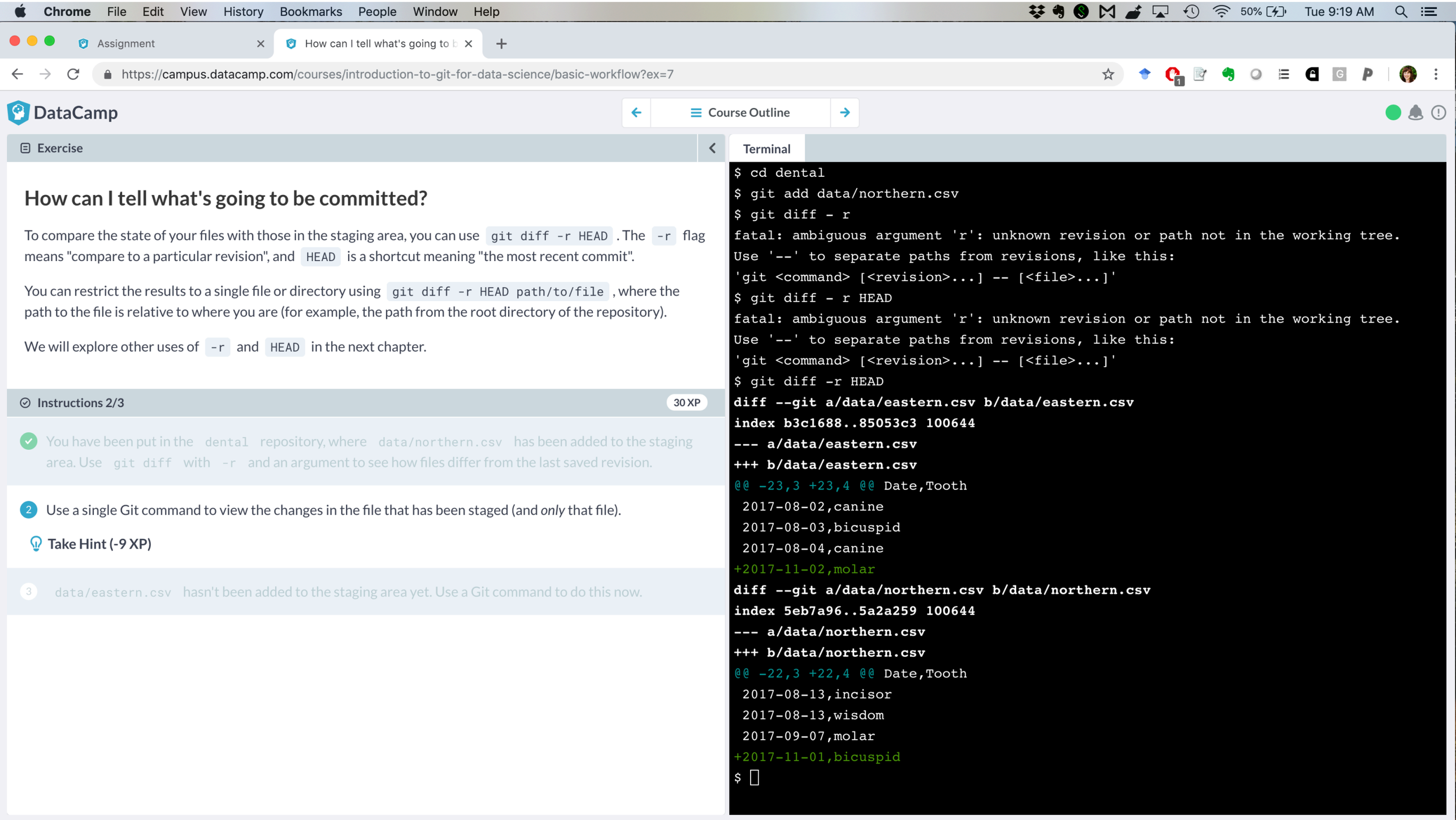
DataCamp for Stat7350



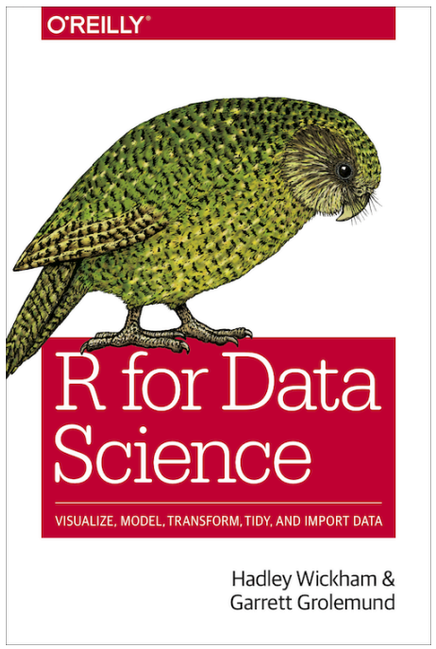
DataCamp for Stat7350

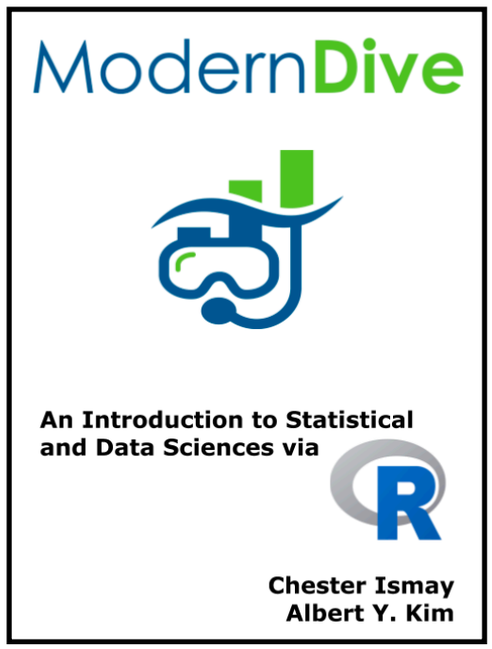


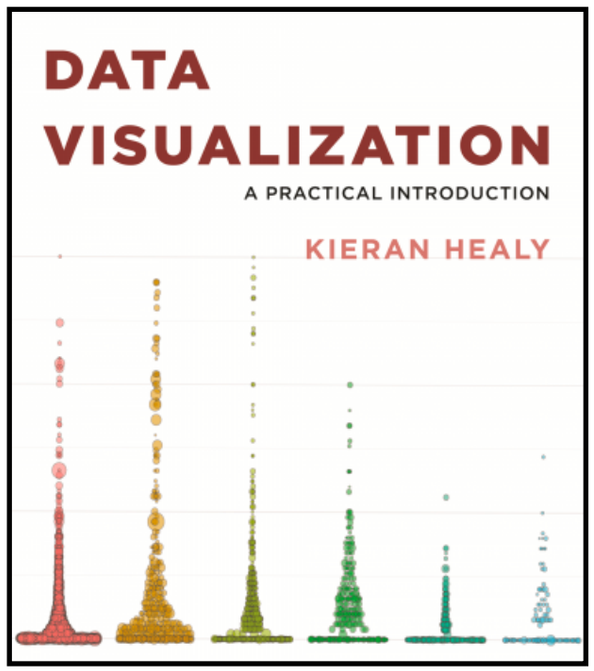
DataCamp for Stat7350

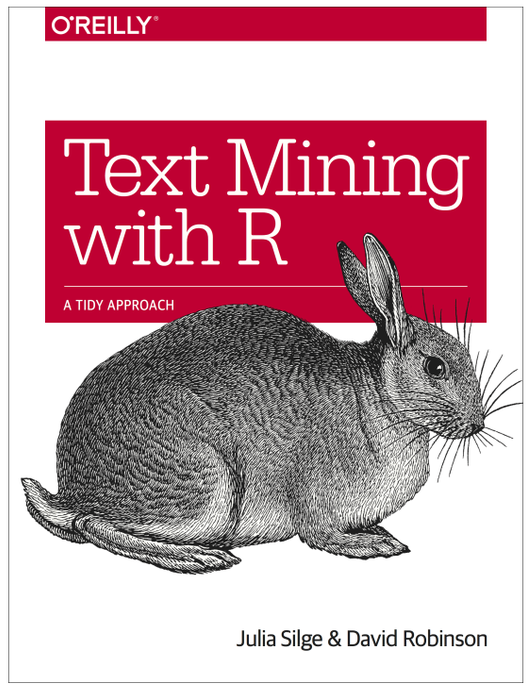


(a sampling of) # RResources





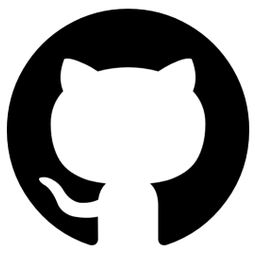




What They Forgot to Teach You

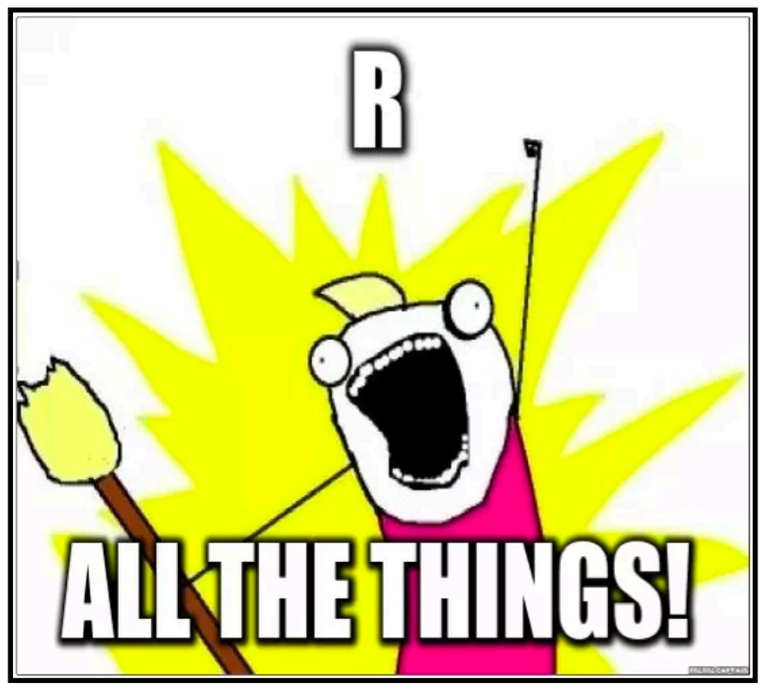
Happy gitR

Advanced Statistical Computing



R Courses

Tidy Tuesday



http://github.com/acgerstein/seminars/RAllTheThings-AleezaGerstein.pdf