YingHu\_assignment4

Ying Hu

9/16/2019

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

?UCBAdmissions

## starting httpd help server ... done

str(UCBAdmissions)

## 'table' num [1:2, 1:2, 1:6] 512 313 89 19 353 207 17 8 120 205 ...  
## - attr(\*, "dimnames")=List of 3  
## ..$ Admit : chr [1:2] "Admitted" "Rejected"  
## ..$ Gender: chr [1:2] "Male" "Female"  
## ..$ Dept : chr [1:6] "A" "B" "C" "D" ...

typeof(UCBAdmissions)

## [1] "double"

head(UCBAdmissions)

## [1] 512 313 89 19 353 207

class(UCBAdmissions)

## [1] "table"

is.array((UCBAdmissions))

## [1] TRUE

is.list(UCBAdmissions)

## [1] FALSE

dimnames(UCBAdmissions)

## $Admit  
## [1] "Admitted" "Rejected"  
##   
## $Gender  
## [1] "Male" "Female"  
##   
## $Dept  
## [1] "A" "B" "C" "D" "E" "F"

dimnames(UCBAdmissions)[3]

## $Dept  
## [1] "A" "B" "C" "D" "E" "F"

margin.table(UCBAdmissions,1)

## Admit  
## Admitted Rejected   
## 1755 2771

margin.table(UCBAdmissions,2)

## Gender  
## Male Female   
## 2691 1835

margin.table(UCBAdmissions,3)

## Dept  
## A B C D E F   
## 933 585 918 792 584 714

margin.table(UCBAdmissions)

## [1] 4526

?margin.table  
  
UCBtotal <- apply(UCBAdmissions, c(2,1), sum) #c(1,2) maybe more natural but this way it matches my ppt  
UCBtotal

## Admit  
## Gender Admitted Rejected  
## Male 1198 1493  
## Female 557 1278

apply(UCBAdmissions, 3, function(x) (x[1,1]\*x[2,2])/(x[1,2]\*x[2,1])) ###Transition back to ppt

## A B C D E F   
## 0.3492120 0.8025007 1.1330596 0.9212838 1.2216312 0.8278727

## Including Plots

You can also embed plots, for example:

m<-matrix(data=cbind(rnorm(30,0), rnorm(30,2), rnorm(30,5)), nrow=30, ncol=3)  
head(m)#rnorm(n, mean = , sd = ) combine the data as matrix with 30rows and 3 columns of mean which is 0,2,5

## [,1] [,2] [,3]  
## [1,] 0.34661830 1.0370805 4.975816  
## [2,] -0.11062471 1.8766144 3.755149  
## [3,] 0.80583314 2.7794288 4.802954  
## [4,] 0.07846576 0.9802475 5.123994  
## [5,] -0.44834612 2.5122526 4.204726  
## [6,] 1.15771573 2.1607918 4.108478

apply(m,1,mean)#mean of the rows,dimension1 or margin1

## [1] 2.119838 1.840380 2.796072 2.060903 2.089544 2.475662 2.177212  
## [8] 3.296007 1.865925 1.603201 2.091604 1.806420 2.620479 1.363886  
## [15] 1.677064 2.422284 3.132312 2.239034 2.453523 2.499076 3.098872  
## [22] 1.585591 2.557172 2.226101 2.304749 3.350385 2.130009 2.098708  
## [29] 3.427427 2.422288

apply(m,2,mean)#mean of the column, dimension2, or margin2

## [1] 0.07524092 1.85617523 5.05175654

#if we want to look at only at positive numbers  
#how many negative numbers we have?  
  
apply(m,2,function(x) length(x[x<0]))# output [1] 16 0 0--16 values in column 1 are negative, 2 & 5 column is 0.

## [1] 15 1 0

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#sapply and lapply  
sapply(1:3, function(x) x^2)#output[1] 1 4 9,shown as index value

## [1] 1 4 9

lapply(1:3, function(x) x^2)# output shown as vetors

## [[1]]  
## [1] 1  
##   
## [[2]]  
## [1] 4  
##   
## [[3]]  
## [1] 9

sapply(1:3, function(x)x^2,simplify=F)

## [[1]]  
## [1] 1  
##   
## [[2]]  
## [1] 4  
##   
## [[3]]  
## [1] 9

#function arguments  
#Argument Matching (1)  
mydata<-rnorm(100)#set as an object  
sd(mydata) #funtion standard deviation of the values

## [1] 1.039585

sd(x=mydata)#set a vector

## [1] 1.039585

sd(x=mydata, na.rm=FALSE)#logical. Should missing values be removed?

## [1] 1.039585

sd(na.rm=FALSE,x=mydata)

## [1] 1.039585

sd(na.rm=FALSE, mydata)

## [1] 1.039585

#Argument Matching(2)  
#args(lm)  
#function(formula, data, subset, weights, na.action, method="qr",  
 #model=TRUE, x=FALSE, y=FALSE, qr=TRUE, singular.ok=TRUE, contrasts=NULL, offset, ...)  
  
#the following two calls are equivalent.  
#lm(data=mydata, y~x, model=FALSE, 1:100)  
#lm(y~x, mydata, 1:100, model=FALSE)

#Defining a Function  
#f<-function(a, b=1, c=2, d=NULL){return a}  
#In addition to not specifying a default value, you can also set an argument value to NULL  
f<- function(a, b=20, c=200, d=NULL)list(a,b,c,d)#out put is shown as a list.   
f(2)

## [[1]]  
## [1] 2  
##   
## [[2]]  
## [1] 20  
##   
## [[3]]  
## [1] 200  
##   
## [[4]]  
## NULL

#Lazy Evaluation  
f<-function(a,b){a^2}#two arguments but only call a not b since a requests square root  
f(2)

## [1] 4

#Lazy Evaluation(2)  
f<- function(a,b){print(a);print(b)}#b did not have to be evaluated until after print(a)  
#f(89)

#The"..."Argument(1)  
myplot<- function(x, y, type="I",...){plot(x,y,type=type,...)}#"..."extending another functions  
args(paste)

## function (..., sep = " ", collapse = NULL)   
## NULL

function (..., sep = " ", collapse = NULL)  
 args(cat)

## function (..., sep = " ", collapse = NULL)  
## args(cat)

function (..., file = "", sep = " ", fill = FALSE, labels = NULL,   
 append = FALSE)  
paste("a", "b", sep=":")#output [1] "a:b"

## function (..., file = "", sep = " ", fill = FALSE, labels = NULL,   
## append = FALSE)  
## paste("a", "b", sep=":")

paste("a","b", se=":")#output [1] "a b :"

## [1] "a b :"

#Lexical Scoping  
f<-function(x,y){x^2+y/z}#x & Y are local variables  
z<-5 #Z is free variable  
f(1,2)#output [1] 1.4

## [1] 1.4

make.power<-function(n){ #function(n)include another function/return pow  
 pow<- function(x){  
 x^n  
 }  
 pow  
}  
  
cube<-make.power(3)#n=3  
square<-make.power(2)#n=2  
cube(3)#output [1] 27 n^3

## [1] 27

square(3)#output [1] 9 n^2

## [1] 9

#Function closure  
#Function enviroment  
ls(environment(cube))#output [1] "n" "pow"

## [1] "n" "pow"

get("n", environment(cube))#output [1] 3

## [1] 3

ls(environment(square))#output [1] "n" "pow"

## [1] "n" "pow"

get("n", environment(square))#output [1] 2

## [1] 2

#Local vs. Dynamic scoping(1)  
y<-10  
f<-function(x){y<-2; y^2+g(x)}#x will be the arguments f(3)=2^2+3\*10=34  
g<-function(x){x\*y}#g(3)=3\*10=30  
f(3)

## [1] 34

g(3)

## [1] 30

g<-function(x) {a<-3; x+a+y}  
#g(2)#Error in g(2): object"y" not found  
   
y<- 3   
g(2)#output 8

## [1] 8

#2 Collect data from Google trends since 2004 for the terms"beer" and "wine", and develop visualization of the correcspooding trends using ggplot.  
  
install.packages("tidyverse",repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/wingy/OneDrive/Documents/R/win-library/3.6'  
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\wingy\AppData\Local\Temp\RtmpKo1c4c\downloaded\_packages

data(package="tidyverse")

## no data sets found

install.packages("ggplot2",repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/wingy/OneDrive/Documents/R/win-library/3.6'  
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\wingy\AppData\Local\Temp\RtmpKo1c4c\downloaded\_packages

data(package="ggplot2")

x1 <- read.csv("C:/Users/wingy/Desktop/multiTimeline.csv", header=F)  
### !  
x <- read.csv("C:/Users/wingy/Desktop/multiTimeline.csv", header=F)  
x

## V1 V2 V3  
## 1 Category: All categories   
## 2 Month beer: (United States) wine: (United States)  
## 3 2004-01 33 61  
## 4 2004-02 32 55  
## 5 2004-03 34 53  
## 6 2004-04 35 53  
## 7 2004-05 33 53  
## 8 2004-06 35 57  
## 9 2004-07 36 57  
## 10 2004-08 36 58  
## 11 2004-09 34 56  
## 12 2004-10 33 57  
## 13 2004-11 33 67  
## 14 2004-12 34 88  
## 15 2005-01 29 61  
## 16 2005-02 32 62  
## 17 2005-03 32 57  
## 18 2005-04 31 57  
## 19 2005-05 33 61  
## 20 2005-06 35 60  
## 21 2005-07 36 62  
## 22 2005-08 35 63  
## 23 2005-09 36 62  
## 24 2005-10 37 62  
## 25 2005-11 34 72  
## 26 2005-12 37 91  
## 27 2006-01 31 64  
## 28 2006-02 33 61  
## 29 2006-03 34 58  
## 30 2006-04 33 58  
## 31 2006-05 35 59  
## 32 2006-06 38 61  
## 33 2006-07 38 61  
## 34 2006-08 38 62  
## 35 2006-09 41 64  
## 36 2006-10 35 63  
## 37 2006-11 33 72  
## 38 2006-12 39 93  
## 39 2007-01 34 63  
## 40 2007-02 34 62  
## 41 2007-03 36 59  
## 42 2007-04 33 58  
## 43 2007-05 35 58  
## 44 2007-06 39 63  
## 45 2007-07 38 63  
## 46 2007-08 38 61  
## 47 2007-09 38 63  
## 48 2007-10 38 62  
## 49 2007-11 35 71  
## 50 2007-12 37 87  
## 51 2008-01 33 61  
## 52 2008-02 33 61  
## 53 2008-03 35 57  
## 54 2008-04 35 58  
## 55 2008-05 36 59  
## 56 2008-06 39 59  
## 57 2008-07 40 58  
## 58 2008-08 38 57  
## 59 2008-09 37 58  
## 60 2008-10 36 57  
## 61 2008-11 33 65  
## 62 2008-12 36 80  
## 63 2009-01 34 59  
## 64 2009-02 35 59  
## 65 2009-03 36 54  
## 66 2009-04 36 55  
## 67 2009-05 37 56  
## 68 2009-06 39 56  
## 69 2009-07 43 56  
## 70 2009-08 43 58  
## 71 2009-09 44 59  
## 72 2009-10 40 57  
## 73 2009-11 36 62  
## 74 2009-12 39 77  
## 75 2010-01 36 57  
## 76 2010-02 37 56  
## 77 2010-03 37 54  
## 78 2010-04 36 54  
## 79 2010-05 38 55  
## 80 2010-06 41 54  
## 81 2010-07 43 55  
## 82 2010-08 42 57  
## 83 2010-09 41 58  
## 84 2010-10 39 56  
## 85 2010-11 36 62  
## 86 2010-12 40 76  
## 87 2011-01 44 65  
## 88 2011-02 43 67  
## 89 2011-03 44 59  
## 90 2011-04 44 62  
## 91 2011-05 47 62  
## 92 2011-06 51 62  
## 93 2011-07 52 64  
## 94 2011-08 47 63  
## 95 2011-09 48 65  
## 96 2011-10 47 65  
## 97 2011-11 44 70  
## 98 2011-12 48 84  
## 99 2012-01 44 65  
## 100 2012-02 46 64  
## 101 2012-03 47 60  
## 102 2012-04 47 61  
## 103 2012-05 52 63  
## 104 2012-06 55 63  
## 105 2012-07 54 63  
## 106 2012-08 52 64  
## 107 2012-09 53 66  
## 108 2012-10 47 61  
## 109 2012-11 44 68  
## 110 2012-12 49 84  
## 111 2013-01 44 63  
## 112 2013-02 47 64  
## 113 2013-03 50 62  
## 114 2013-04 47 60  
## 115 2013-05 51 62  
## 116 2013-06 55 62  
## 117 2013-07 55 62  
## 118 2013-08 55 63  
## 119 2013-09 51 62  
## 120 2013-10 49 62  
## 121 2013-11 49 72  
## 122 2013-12 51 83  
## 123 2014-01 48 65  
## 124 2014-02 50 65  
## 125 2014-03 52 61  
## 126 2014-04 51 59  
## 127 2014-05 55 63  
## 128 2014-06 59 61  
## 129 2014-07 62 63  
## 130 2014-08 58 63  
## 131 2014-09 51 63  
## 132 2014-10 49 62  
## 133 2014-11 47 69  
## 134 2014-12 51 83  
## 135 2015-01 48 65  
## 136 2015-02 46 63  
## 137 2015-03 51 63  
## 138 2015-04 53 59  
## 139 2015-05 55 61  
## 140 2015-06 59 61  
## 141 2015-07 62 64  
## 142 2015-08 63 67  
## 143 2015-09 57 65  
## 144 2015-10 52 65  
## 145 2015-11 49 73  
## 146 2015-12 51 88  
## 147 2016-01 49 69  
## 148 2016-02 51 69  
## 149 2016-03 54 64  
## 150 2016-04 57 66  
## 151 2016-05 59 69  
## 152 2016-06 62 68  
## 153 2016-07 63 68  
## 154 2016-08 56 66  
## 155 2016-09 56 69  
## 156 2016-10 53 68  
## 157 2016-11 49 75  
## 158 2016-12 53 94  
## 159 2017-01 49 69  
## 160 2017-02 52 69  
## 161 2017-03 55 66  
## 162 2017-04 57 66  
## 163 2017-05 57 70  
## 164 2017-06 66 70  
## 165 2017-07 67 73  
## 166 2017-08 59 69  
## 167 2017-09 57 71  
## 168 2017-10 52 71  
## 169 2017-11 50 83  
## 170 2017-12 55 100  
## 171 2018-01 47 70  
## 172 2018-02 50 70  
## 173 2018-03 54 69  
## 174 2018-04 55 69  
## 175 2018-05 55 72  
## 176 2018-06 63 72  
## 177 2018-07 64 73  
## 178 2018-08 59 73  
## 179 2018-09 56 71  
## 180 2018-10 52 69  
## 181 2018-11 49 79  
## 182 2018-12 55 99  
## 183 2019-01 49 73  
## 184 2019-02 50 70  
## 185 2019-03 54 68  
## 186 2019-04 53 68  
## 187 2019-05 55 75  
## 188 2019-06 61 72  
## 189 2019-07 61 71  
## 190 2019-08 60 71  
## 191 2019-09 54 69

head(x)

## V1 V2 V3  
## 1 Category: All categories   
## 2 Month beer: (United States) wine: (United States)  
## 3 2004-01 33 61  
## 4 2004-02 32 55  
## 5 2004-03 34 53  
## 6 2004-04 35 53

dim(x)

## [1] 191 3

str(x)

## 'data.frame': 191 obs. of 3 variables:  
## $ V1: Factor w/ 191 levels "2004-01","2004-02",..: 190 191 1 2 3 4 5 6 7 8 ...  
## $ V2: Factor w/ 38 levels "","29","31","32",..: 1 38 5 4 6 7 5 7 8 8 ...  
## $ V3: Factor w/ 37 levels "","100","53",..: 1 37 11 5 3 3 3 7 7 8 ...

library(ggplot2)  
  
trends1 <- x[c(-1,-2),]  
trends1

## V1 V2 V3  
## 3 2004-01 33 61  
## 4 2004-02 32 55  
## 5 2004-03 34 53  
## 6 2004-04 35 53  
## 7 2004-05 33 53  
## 8 2004-06 35 57  
## 9 2004-07 36 57  
## 10 2004-08 36 58  
## 11 2004-09 34 56  
## 12 2004-10 33 57  
## 13 2004-11 33 67  
## 14 2004-12 34 88  
## 15 2005-01 29 61  
## 16 2005-02 32 62  
## 17 2005-03 32 57  
## 18 2005-04 31 57  
## 19 2005-05 33 61  
## 20 2005-06 35 60  
## 21 2005-07 36 62  
## 22 2005-08 35 63  
## 23 2005-09 36 62  
## 24 2005-10 37 62  
## 25 2005-11 34 72  
## 26 2005-12 37 91  
## 27 2006-01 31 64  
## 28 2006-02 33 61  
## 29 2006-03 34 58  
## 30 2006-04 33 58  
## 31 2006-05 35 59  
## 32 2006-06 38 61  
## 33 2006-07 38 61  
## 34 2006-08 38 62  
## 35 2006-09 41 64  
## 36 2006-10 35 63  
## 37 2006-11 33 72  
## 38 2006-12 39 93  
## 39 2007-01 34 63  
## 40 2007-02 34 62  
## 41 2007-03 36 59  
## 42 2007-04 33 58  
## 43 2007-05 35 58  
## 44 2007-06 39 63  
## 45 2007-07 38 63  
## 46 2007-08 38 61  
## 47 2007-09 38 63  
## 48 2007-10 38 62  
## 49 2007-11 35 71  
## 50 2007-12 37 87  
## 51 2008-01 33 61  
## 52 2008-02 33 61  
## 53 2008-03 35 57  
## 54 2008-04 35 58  
## 55 2008-05 36 59  
## 56 2008-06 39 59  
## 57 2008-07 40 58  
## 58 2008-08 38 57  
## 59 2008-09 37 58  
## 60 2008-10 36 57  
## 61 2008-11 33 65  
## 62 2008-12 36 80  
## 63 2009-01 34 59  
## 64 2009-02 35 59  
## 65 2009-03 36 54  
## 66 2009-04 36 55  
## 67 2009-05 37 56  
## 68 2009-06 39 56  
## 69 2009-07 43 56  
## 70 2009-08 43 58  
## 71 2009-09 44 59  
## 72 2009-10 40 57  
## 73 2009-11 36 62  
## 74 2009-12 39 77  
## 75 2010-01 36 57  
## 76 2010-02 37 56  
## 77 2010-03 37 54  
## 78 2010-04 36 54  
## 79 2010-05 38 55  
## 80 2010-06 41 54  
## 81 2010-07 43 55  
## 82 2010-08 42 57  
## 83 2010-09 41 58  
## 84 2010-10 39 56  
## 85 2010-11 36 62  
## 86 2010-12 40 76  
## 87 2011-01 44 65  
## 88 2011-02 43 67  
## 89 2011-03 44 59  
## 90 2011-04 44 62  
## 91 2011-05 47 62  
## 92 2011-06 51 62  
## 93 2011-07 52 64  
## 94 2011-08 47 63  
## 95 2011-09 48 65  
## 96 2011-10 47 65  
## 97 2011-11 44 70  
## 98 2011-12 48 84  
## 99 2012-01 44 65  
## 100 2012-02 46 64  
## 101 2012-03 47 60  
## 102 2012-04 47 61  
## 103 2012-05 52 63  
## 104 2012-06 55 63  
## 105 2012-07 54 63  
## 106 2012-08 52 64  
## 107 2012-09 53 66  
## 108 2012-10 47 61  
## 109 2012-11 44 68  
## 110 2012-12 49 84  
## 111 2013-01 44 63  
## 112 2013-02 47 64  
## 113 2013-03 50 62  
## 114 2013-04 47 60  
## 115 2013-05 51 62  
## 116 2013-06 55 62  
## 117 2013-07 55 62  
## 118 2013-08 55 63  
## 119 2013-09 51 62  
## 120 2013-10 49 62  
## 121 2013-11 49 72  
## 122 2013-12 51 83  
## 123 2014-01 48 65  
## 124 2014-02 50 65  
## 125 2014-03 52 61  
## 126 2014-04 51 59  
## 127 2014-05 55 63  
## 128 2014-06 59 61  
## 129 2014-07 62 63  
## 130 2014-08 58 63  
## 131 2014-09 51 63  
## 132 2014-10 49 62  
## 133 2014-11 47 69  
## 134 2014-12 51 83  
## 135 2015-01 48 65  
## 136 2015-02 46 63  
## 137 2015-03 51 63  
## 138 2015-04 53 59  
## 139 2015-05 55 61  
## 140 2015-06 59 61  
## 141 2015-07 62 64  
## 142 2015-08 63 67  
## 143 2015-09 57 65  
## 144 2015-10 52 65  
## 145 2015-11 49 73  
## 146 2015-12 51 88  
## 147 2016-01 49 69  
## 148 2016-02 51 69  
## 149 2016-03 54 64  
## 150 2016-04 57 66  
## 151 2016-05 59 69  
## 152 2016-06 62 68  
## 153 2016-07 63 68  
## 154 2016-08 56 66  
## 155 2016-09 56 69  
## 156 2016-10 53 68  
## 157 2016-11 49 75  
## 158 2016-12 53 94  
## 159 2017-01 49 69  
## 160 2017-02 52 69  
## 161 2017-03 55 66  
## 162 2017-04 57 66  
## 163 2017-05 57 70  
## 164 2017-06 66 70  
## 165 2017-07 67 73  
## 166 2017-08 59 69  
## 167 2017-09 57 71  
## 168 2017-10 52 71  
## 169 2017-11 50 83  
## 170 2017-12 55 100  
## 171 2018-01 47 70  
## 172 2018-02 50 70  
## 173 2018-03 54 69  
## 174 2018-04 55 69  
## 175 2018-05 55 72  
## 176 2018-06 63 72  
## 177 2018-07 64 73  
## 178 2018-08 59 73  
## 179 2018-09 56 71  
## 180 2018-10 52 69  
## 181 2018-11 49 79  
## 182 2018-12 55 99  
## 183 2019-01 49 73  
## 184 2019-02 50 70  
## 185 2019-03 54 68  
## 186 2019-04 53 68  
## 187 2019-05 55 75  
## 188 2019-06 61 72  
## 189 2019-07 61 71  
## 190 2019-08 60 71  
## 191 2019-09 54 69

colnames(trends1) <- c("month", "beer", "wine")#set 3 columns  
trends1

## month beer wine  
## 3 2004-01 33 61  
## 4 2004-02 32 55  
## 5 2004-03 34 53  
## 6 2004-04 35 53  
## 7 2004-05 33 53  
## 8 2004-06 35 57  
## 9 2004-07 36 57  
## 10 2004-08 36 58  
## 11 2004-09 34 56  
## 12 2004-10 33 57  
## 13 2004-11 33 67  
## 14 2004-12 34 88  
## 15 2005-01 29 61  
## 16 2005-02 32 62  
## 17 2005-03 32 57  
## 18 2005-04 31 57  
## 19 2005-05 33 61  
## 20 2005-06 35 60  
## 21 2005-07 36 62  
## 22 2005-08 35 63  
## 23 2005-09 36 62  
## 24 2005-10 37 62  
## 25 2005-11 34 72  
## 26 2005-12 37 91  
## 27 2006-01 31 64  
## 28 2006-02 33 61  
## 29 2006-03 34 58  
## 30 2006-04 33 58  
## 31 2006-05 35 59  
## 32 2006-06 38 61  
## 33 2006-07 38 61  
## 34 2006-08 38 62  
## 35 2006-09 41 64  
## 36 2006-10 35 63  
## 37 2006-11 33 72  
## 38 2006-12 39 93  
## 39 2007-01 34 63  
## 40 2007-02 34 62  
## 41 2007-03 36 59  
## 42 2007-04 33 58  
## 43 2007-05 35 58  
## 44 2007-06 39 63  
## 45 2007-07 38 63  
## 46 2007-08 38 61  
## 47 2007-09 38 63  
## 48 2007-10 38 62  
## 49 2007-11 35 71  
## 50 2007-12 37 87  
## 51 2008-01 33 61  
## 52 2008-02 33 61  
## 53 2008-03 35 57  
## 54 2008-04 35 58  
## 55 2008-05 36 59  
## 56 2008-06 39 59  
## 57 2008-07 40 58  
## 58 2008-08 38 57  
## 59 2008-09 37 58  
## 60 2008-10 36 57  
## 61 2008-11 33 65  
## 62 2008-12 36 80  
## 63 2009-01 34 59  
## 64 2009-02 35 59  
## 65 2009-03 36 54  
## 66 2009-04 36 55  
## 67 2009-05 37 56  
## 68 2009-06 39 56  
## 69 2009-07 43 56  
## 70 2009-08 43 58  
## 71 2009-09 44 59  
## 72 2009-10 40 57  
## 73 2009-11 36 62  
## 74 2009-12 39 77  
## 75 2010-01 36 57  
## 76 2010-02 37 56  
## 77 2010-03 37 54  
## 78 2010-04 36 54  
## 79 2010-05 38 55  
## 80 2010-06 41 54  
## 81 2010-07 43 55  
## 82 2010-08 42 57  
## 83 2010-09 41 58  
## 84 2010-10 39 56  
## 85 2010-11 36 62  
## 86 2010-12 40 76  
## 87 2011-01 44 65  
## 88 2011-02 43 67  
## 89 2011-03 44 59  
## 90 2011-04 44 62  
## 91 2011-05 47 62  
## 92 2011-06 51 62  
## 93 2011-07 52 64  
## 94 2011-08 47 63  
## 95 2011-09 48 65  
## 96 2011-10 47 65  
## 97 2011-11 44 70  
## 98 2011-12 48 84  
## 99 2012-01 44 65  
## 100 2012-02 46 64  
## 101 2012-03 47 60  
## 102 2012-04 47 61  
## 103 2012-05 52 63  
## 104 2012-06 55 63  
## 105 2012-07 54 63  
## 106 2012-08 52 64  
## 107 2012-09 53 66  
## 108 2012-10 47 61  
## 109 2012-11 44 68  
## 110 2012-12 49 84  
## 111 2013-01 44 63  
## 112 2013-02 47 64  
## 113 2013-03 50 62  
## 114 2013-04 47 60  
## 115 2013-05 51 62  
## 116 2013-06 55 62  
## 117 2013-07 55 62  
## 118 2013-08 55 63  
## 119 2013-09 51 62  
## 120 2013-10 49 62  
## 121 2013-11 49 72  
## 122 2013-12 51 83  
## 123 2014-01 48 65  
## 124 2014-02 50 65  
## 125 2014-03 52 61  
## 126 2014-04 51 59  
## 127 2014-05 55 63  
## 128 2014-06 59 61  
## 129 2014-07 62 63  
## 130 2014-08 58 63  
## 131 2014-09 51 63  
## 132 2014-10 49 62  
## 133 2014-11 47 69  
## 134 2014-12 51 83  
## 135 2015-01 48 65  
## 136 2015-02 46 63  
## 137 2015-03 51 63  
## 138 2015-04 53 59  
## 139 2015-05 55 61  
## 140 2015-06 59 61  
## 141 2015-07 62 64  
## 142 2015-08 63 67  
## 143 2015-09 57 65  
## 144 2015-10 52 65  
## 145 2015-11 49 73  
## 146 2015-12 51 88  
## 147 2016-01 49 69  
## 148 2016-02 51 69  
## 149 2016-03 54 64  
## 150 2016-04 57 66  
## 151 2016-05 59 69  
## 152 2016-06 62 68  
## 153 2016-07 63 68  
## 154 2016-08 56 66  
## 155 2016-09 56 69  
## 156 2016-10 53 68  
## 157 2016-11 49 75  
## 158 2016-12 53 94  
## 159 2017-01 49 69  
## 160 2017-02 52 69  
## 161 2017-03 55 66  
## 162 2017-04 57 66  
## 163 2017-05 57 70  
## 164 2017-06 66 70  
## 165 2017-07 67 73  
## 166 2017-08 59 69  
## 167 2017-09 57 71  
## 168 2017-10 52 71  
## 169 2017-11 50 83  
## 170 2017-12 55 100  
## 171 2018-01 47 70  
## 172 2018-02 50 70  
## 173 2018-03 54 69  
## 174 2018-04 55 69  
## 175 2018-05 55 72  
## 176 2018-06 63 72  
## 177 2018-07 64 73  
## 178 2018-08 59 73  
## 179 2018-09 56 71  
## 180 2018-10 52 69  
## 181 2018-11 49 79  
## 182 2018-12 55 99  
## 183 2019-01 49 73  
## 184 2019-02 50 70  
## 185 2019-03 54 68  
## 186 2019-04 53 68  
## 187 2019-05 55 75  
## 188 2019-06 61 72  
## 189 2019-07 61 71  
## 190 2019-08 60 71  
## 191 2019-09 54 69

View(trends1)  
dim(trends1)

## [1] 189 3

str(trends1)

## 'data.frame': 189 obs. of 3 variables:  
## $ month: Factor w/ 191 levels "2004-01","2004-02",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ beer : Factor w/ 38 levels "","29","31","32",..: 5 4 6 7 5 7 8 8 6 5 ...  
## $ wine : Factor w/ 37 levels "","100","53",..: 11 5 3 3 3 7 7 8 6 7 ...

trends1$month <- as.Date(paste(trends1$month,"-01",sep=""))  
head(trends1)

## month beer wine  
## 3 2004-01-01 33 61  
## 4 2004-02-01 32 55  
## 5 2004-03-01 34 53  
## 6 2004-04-01 35 53  
## 7 2004-05-01 33 53  
## 8 2004-06-01 35 57

is.list(trends1)

## [1] TRUE

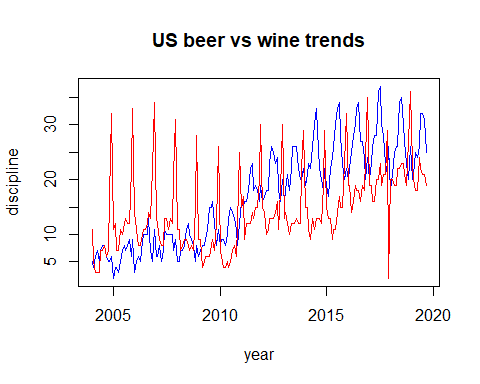
trnd1 <- as.data.frame(lapply(trends1[2:3], FUN = as.numeric))  
trnd2 <- cbind(trends1$month,trnd1)  
colnames(trnd2)[1] <- "month"  
str(trnd2)

## 'data.frame': 189 obs. of 3 variables:  
## $ month: Date, format: "2004-01-01" "2004-02-01" ...  
## $ beer : num 5 4 6 7 5 7 8 8 6 5 ...  
## $ wine : num 11 5 3 3 3 7 7 8 6 7 ...

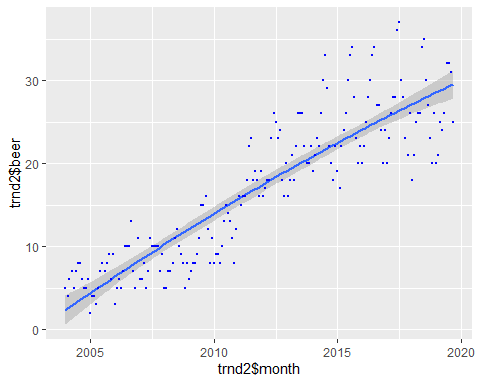
head(trnd2)

## month beer wine  
## 1 2004-01-01 5 11  
## 2 2004-02-01 4 5  
## 3 2004-03-01 6 3  
## 4 2004-04-01 7 3  
## 5 2004-05-01 5 3  
## 6 2004-06-01 7 7

plot(trnd2$month, trnd2$beer, type ="l", main="US beer vs wine trends", col="blue", ylab = "discipline", xlab="year")  
lines(trnd2$month, trnd2$wine, type = "l", col ="red")



library(ggplot2)  
ggplot(trnd2, aes(x=trnd2$month, y=trnd2$beer)) + stat\_smooth(aes(y=trnd2$beer), method=lm, formula = y ~ poly(x,2), level=0.95) + geom\_point(color="blue", size=0.2)



a <- readLines("C:/Users/wingy/Desktop/multiTimeline.csv")  
dim(x)

## [1] 191 3

str(a)

## chr [1:192] "Category: All categories" "" ...

class(a)

## [1] "character"

length(a)

## [1] 192

is.vector(a)

## [1] TRUE

head(a)

## [1] "Category: All categories"   
## [2] ""   
## [3] "Month,beer: (United States),wine: (United States)"  
## [4] "2004-01,33,61"   
## [5] "2004-02,32,55"   
## [6] "2004-03,34,53"

library(tidyverse)

## -- Attaching packages ------------------------------------------------------- tidyverse 1.2.1 --

## v tibble 2.1.3 v purrr 0.3.2  
## v tidyr 0.8.3 v dplyr 0.8.3  
## v readr 1.3.1 v stringr 1.4.0  
## v tibble 2.1.3 v forcats 0.4.0

## -- Conflicts ---------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

dt\_in2 <- read\_csv(a, skip = 2)  
head(dt\_in2)

## # A tibble: 6 x 3  
## Month `beer: (United States)` `wine: (United States)`  
## <chr> <dbl> <dbl>  
## 1 2004-01 33 61  
## 2 2004-02 32 55  
## 3 2004-03 34 53  
## 4 2004-04 35 53  
## 5 2004-05 33 53  
## 6 2004-06 35 57

dtin <- dt\_in2 %>% mutate(Month = lubridate::ymd(Month, truncated = 1))  
head(dtin)

## # A tibble: 6 x 3  
## Month `beer: (United States)` `wine: (United States)`  
## <date> <dbl> <dbl>  
## 1 2004-01-01 33 61  
## 2 2004-02-01 32 55  
## 3 2004-03-01 34 53  
## 4 2004-04-01 35 53  
## 5 2004-05-01 33 53  
## 6 2004-06-01 35 57

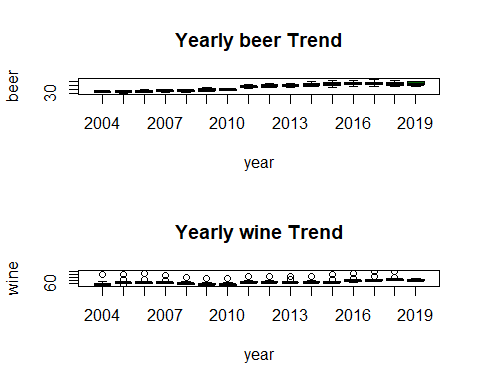
#Need to get years first  
dtin$year <- substring(dtin$Month,1,4)  
dim(dtin)

## [1] 189 4

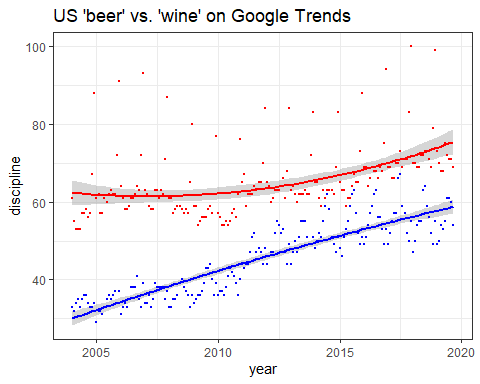
head(dtin)

## # A tibble: 6 x 4  
## Month `beer: (United States)` `wine: (United States)` year   
## <date> <dbl> <dbl> <chr>  
## 1 2004-01-01 33 61 2004   
## 2 2004-02-01 32 55 2004   
## 3 2004-03-01 34 53 2004   
## 4 2004-04-01 35 53 2004   
## 5 2004-05-01 33 53 2004   
## 6 2004-06-01 35 57 2004

par(mfrow = c(2,1))  
boxplot(dtin$`beer: (United States)`~dtin$year, data = dtin, notch = F, col=(c("gold", "darkgreen")),  
 main = "Yearly beer Trend", xlab = "year", ylab = "beer")  
boxplot(dtin$`wine: (United States)`~dtin$year, data = dtin, notch = F, col=(c("gold", "darkgreen")),  
 main = "Yearly wine Trend", xlab = "year", ylab = "wine")



ggplot(dtin, aes(x=dtin$Month, y=dtin$`beer: (United States)`)) +   
 stat\_smooth(aes(y=dtin$`beer: (United States)`, col = "blue"), method=lm, formula = y ~ poly(x,2), level=0.95) +   
  
 stat\_smooth(aes(y = dtin$`wine: (United States)`, col = "red"), method=lm, formula = y ~ poly(x,2), level=0.95) +  
 geom\_point (aes(y = dtin$`beer: (United States)`, col = "blue"), size=0.2) +  
 geom\_point (aes(y = dtin$`wine: (United States)`, col ="red"), size=0.2) +  
  
 scale\_color\_manual(name ="Search Terms", breaks = c("beer", "wine"), values = c("blue","red")) +  
 theme\_bw() +   
 xlab("year") +  
 ylab("discipline") +  
 ggtitle("US 'beer' vs. 'wine' on Google Trends")



### another attempt

```