

# Risk Using Hands on R

# Learning Objectives

- ▶ R Basic Basics
- ▶ FAIR
- ▶ Probability
- ▶ Doing Risk in R
- ▶ Basic Graphics with ggplot2



# Before we get started

- ▶ Do you have RStudio up?
  - ▶ Did you have any trouble with the pre-reqs?
- ▶ Type stuff that looks like this...but not actually this.



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins



```
sankey_demo.R x Untitled6* x Internal One.Rmd x Internal Two.Rmd x Internal Three.Rmd x External Actor Breaches Network Thru... x Internal Actor Erroneously Discloses ... x
```

```
1 print("Hello Risky People")
```

Run Source

Environment History

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

R: Vectors Find in Topic

vector {base}

R Documentation

## Vectors

### Description

vector produces a vector of the given length and mode.

as.vector, a generic, attempts to coerce its argument into a vector of mode mode (the default is to coerce to whichever vector mode is most convenient): if the result is atomic all attributes are removed.

is.vector returns TRUE if x is a vector of the specified mode having no attributes other than names. It returns FALSE otherwise.

### Usage

```
vector(mode = "logical", length = 0)
as.vector(x, mode = "any")
is.vector(x, mode = "any")
```

### Arguments

mode character string naming an atomic mode or "list" or "expression" or (except for vector) "any". Currently, is.vector() allows any type (see [typeof](#)) for mode, and when mode is not "any", is.vector(x, mode) is almost the same as `typeof(x) == mode`.

length a non-negative integer specifying the desired length. For a long vector, i.e., `length > .Machine$integer.max`, it has to be of type "double". Supplying an argument of length other than one is an error.

1:26 [Top Level] ▾

Console ~/ ↵

```
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

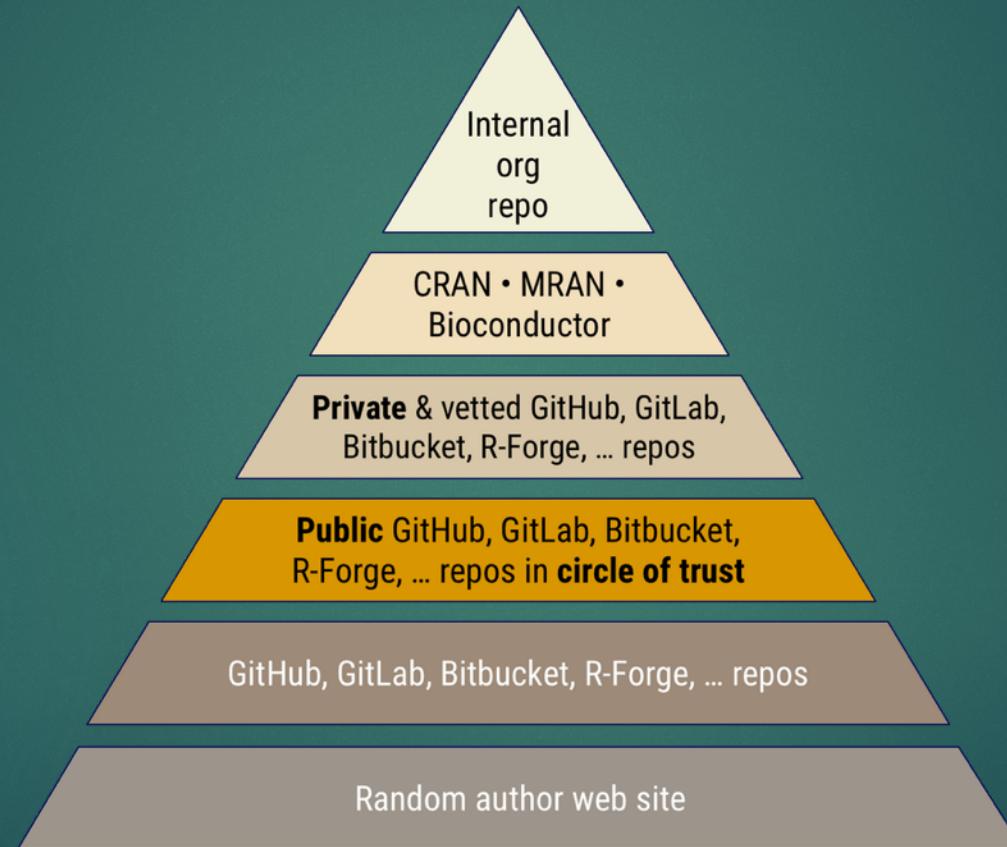
```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
```

```
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.
```

```
> source('~/active-rstudio-document')
[1] "Hello Risky People"
> ?vector
> |
```

# Fantastic Packages And Where to Find Them



@hrbrmstr's (Bob Rudis) HieraRchy of Package Trust

?

YOU CAN TYPE ?<ANY FUNCTION> TO GET HELP IN THE CONSOLE.



ImmaVar <- ImmaVal

THIS IS THE ASSIGNMENT OPERATOR. DON'T LET IT SCARE YOU.

# Vectors

- ▶ Sequenced data elements
- ▶ Share the same type

```
team <- c("Mike", "Justin", "Bo", "Spencer", "Chris")
```

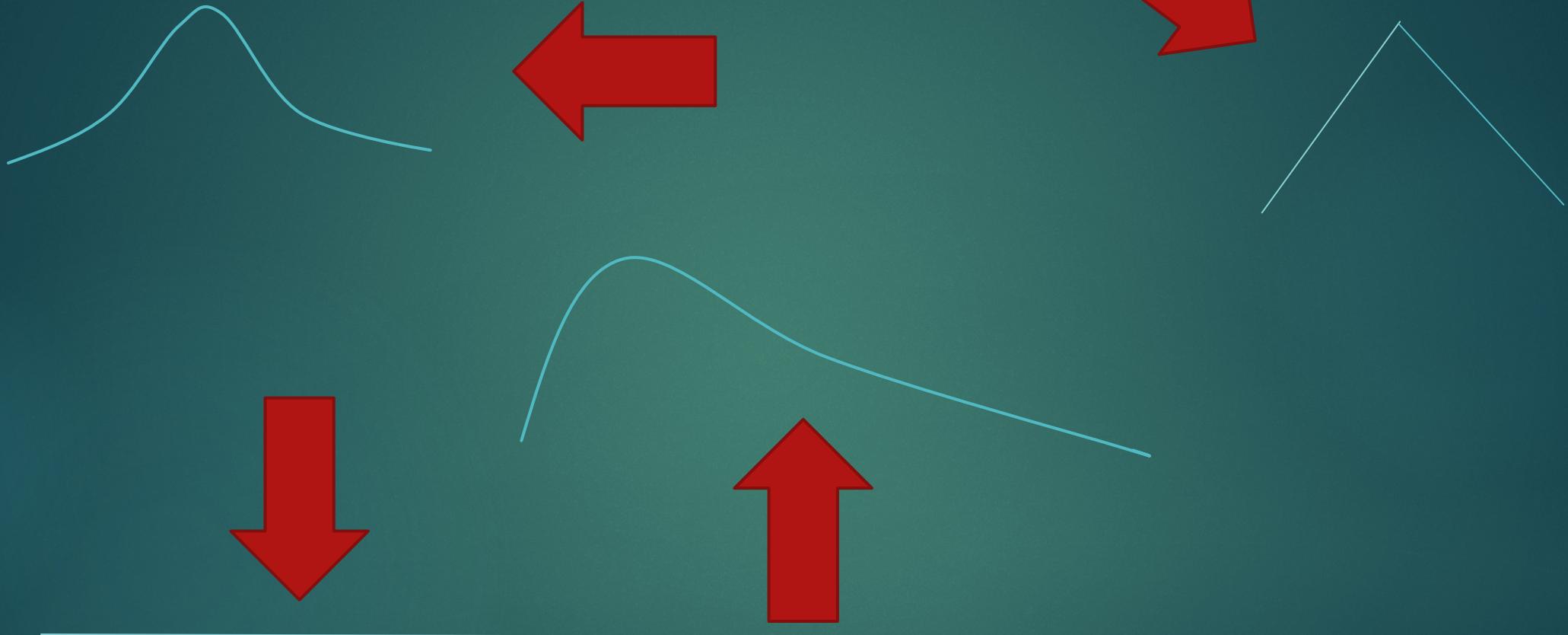
LOAD A VECTOR WITH CHARACTERS



```
team <- team[ !team %in% "Justin" ]
```

REMOVING AN ELEMENT FORM THE TEAM

# Distributions...briefly



```
Ticket_Prices <- runif(50, 1, 10)  
VECTOR OF INTEGERS
```



```
library(modeest)
mlv(Ticket_Prices)
```

MODE ESTIMATION TO THE RESCUE!

# Data Frames

- ▶ Collection of multiple, same-length vectors
- ▶ May contain categorical or numeric values



# PERT

- ▶ Program Evaluation and Review Technique
- ▶ Developed by the U.S. Navy during the 50s
- ▶ Essentially, attempts to balance optimistic, most likely and pessimistic estimates
- ▶ Estimate =  $(o + 4m + p) / 6$

# Beta Distribution

- ▶ Allows for weighted range of probability using knowledge of “hits” and “misses”
- ▶ Let me show you what I mean...



Image of Benintendi redacted

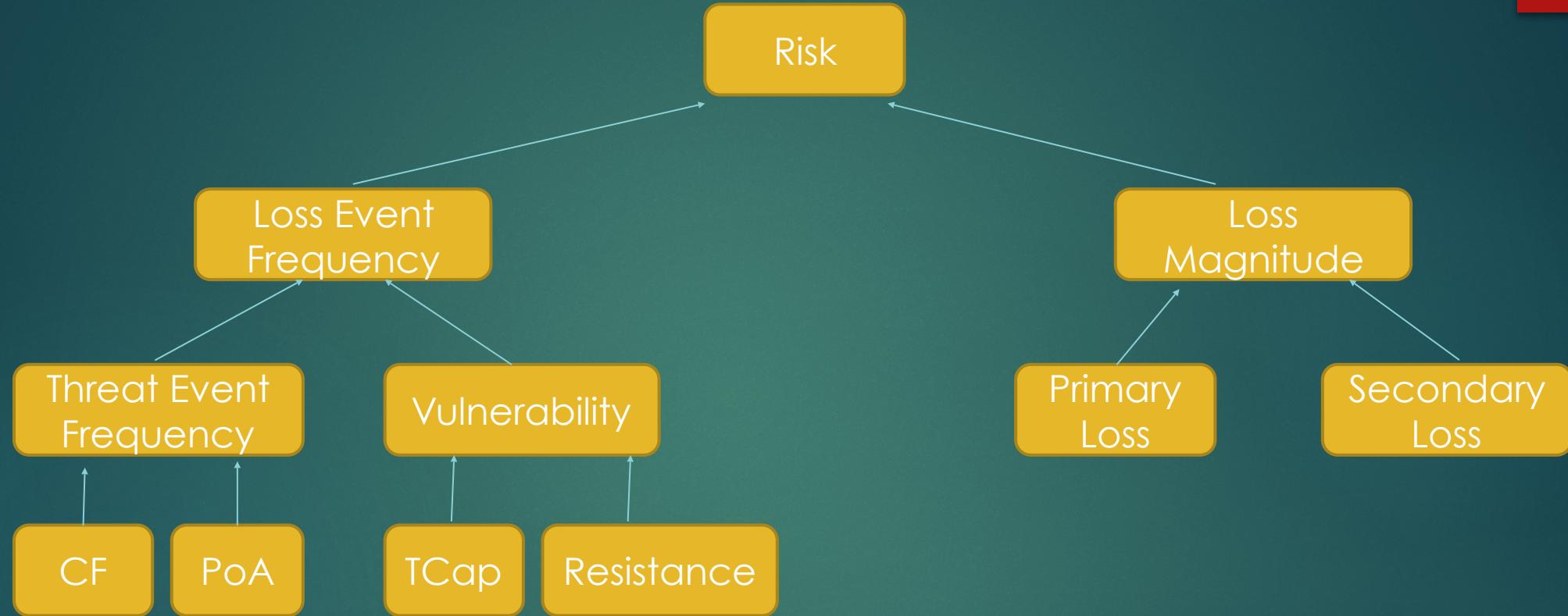
curve( dbeta( x, 154, 565-154 ) )

LET'S SAY RED SOX LF ANDREW BENINTENDI HAS  
154 HITS AND 565 AT BATS

```
bat_avg <- rbeta( 100, 154, 565-154)
```



Welcome to Monte Carlo!  
BREAK



Partial Representation of FAIR Model



```
Some_variable <- rpert(1000, min = 1.5, mode = 2, max = 5, shape  
= 4)
```



## #Workout

You decide to stand up at your desk, walk to your car and start driving.  
Estimate road trip costs trip from <redacted> to the Pocono Mountains using  
Monte Carlo estimation.



```
library(mc2d)

sim.count <- 1000
TEF <- rpert( sim.count, min = 1, mode = 5, max = 10, shape = 4)
vuln <- rpert( sim.count, min = .5, mode = .7, max = .8, shape = 4)
LEF <- TEF * vuln
```

ARE YOU FEELING RISKY YET?

```
mag <- rpert( sim.count, min = 100, mode = 500, max = 10000, shape =  
             4 )  
risk <- LEF * mag
```

MONTE CARLO ANNUAL LOSS EXPOSURE



## #Workout

Analyze the risk of an external threat actor gaining access to confidential information (ePHI) via a misconfigured Azure Blob storage container.

# Plotting Systems

- ▶ Base R
  - ▶ Most basic R plotting system. You build up everything. Sometimes difficult to read.
- ▶ Lattice
  - ▶ More advanced. Takes care of a few things like margins, etc.
- ▶ ggplot2
  - ▶ The best of both worlds!
  - ▶ We're still building up a plot, but we're following a grammar to do it.

```
library (ggplot)
library (scales)

gg <- ggplot(data.frame(risk), aes(x = risk))
print(gg)
```

IMPRESSIVE, NO? NO.

```
gg <- ggplot(data.frame(risk), aes(x = risk))  
gg <- gg + geom_histogram(aes(y = ..count..), color = "deepskyblue4", fill = "#A6761D", binwidth = 10)  
gg <- gg + geom_vline(aes(xintercept=mean(risk, na.rm=T)), color="cadetblue", linetype="solid", size=1)  
gg <- gg + geom_density(fill = "seashell3", alpha = 1/3)  
gg <- gg + scale_x_continuous(labels = dollar)+ xlab("Annual Risk Exposure")  
print(gg)
```

# Modeling



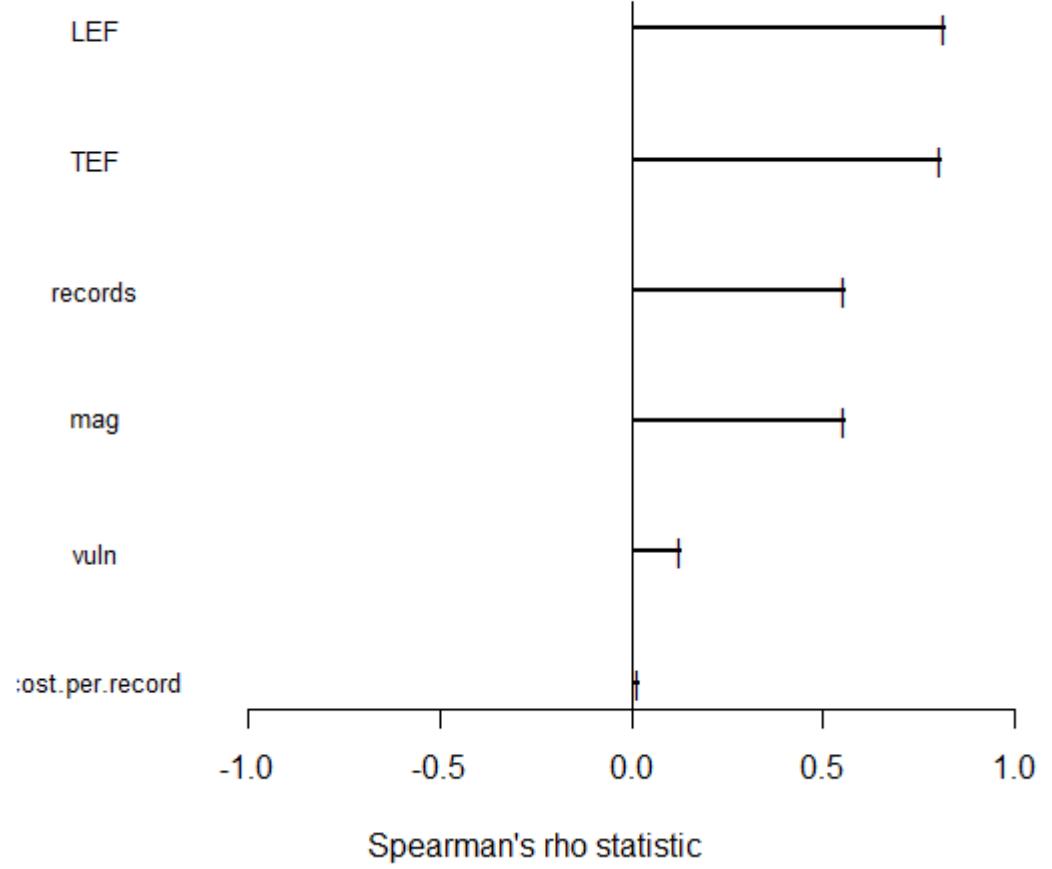
ndvar(2000)

```
TEF <- mcstoc(rpert, min = 1/10, mode = 1/5, max = 1/3, shape = 4)
vuln <- mcstoc(rpert, min = .5, mode = .75, max = .8, shape = 4)
LEF <- TEF * vuln

records <- mcstoc(rpert, min = 10, mode = 50, max = 100, shape = 4)
cost.per.record <- mcstoc(rpert, min = 40, mode = 45, max = 50, shape = 4)
mag <- records * cost.per.record

risk <- LEF * mag

EC1 <- mc(TEF, vuln, LEF, records, cost.per.record, mag, risk)
EC2 <- tornado(EC1)
plot(EC2)
```



# RMarkdown

- ▶ Create reproducible documents
  - ▶ PDF, Docx, epub, HTML...
- ▶ Mix text and code
- ▶ <https://daringfireball.net/projects/markdown/>

# Closing Risk Concepts

- ▶ Unstable
  - ▶ Low TEF
  - ▶ High vulnerability
- ▶ Fragile
  - ▶ High TEF
  - ▶ Single control knocking down vulnerability

# Continue Learning

- ▶ DataCamp – Data Scientist with R <https://www.datacamp.com>
- ▶ Try R <http://tryr.codeschool.com>
- ▶ R Fiddle <http://www.r-fiddle.org>

