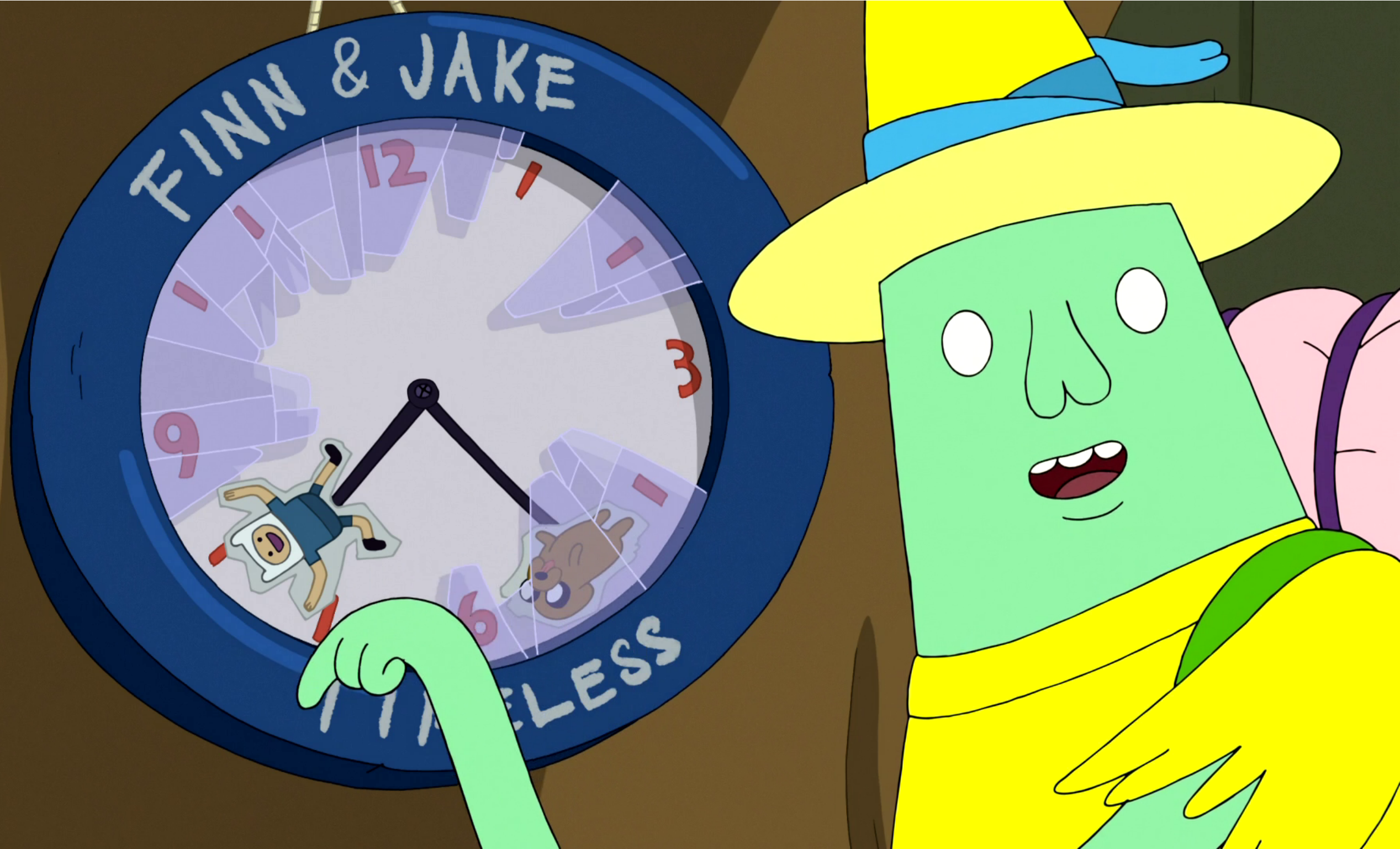


efficient statistics



@cjlortie

assumption: many domain scientists use R
but are not programmers

R is the means to an end



limitation: programming is a separate discipline

efficient **learning**
to code important

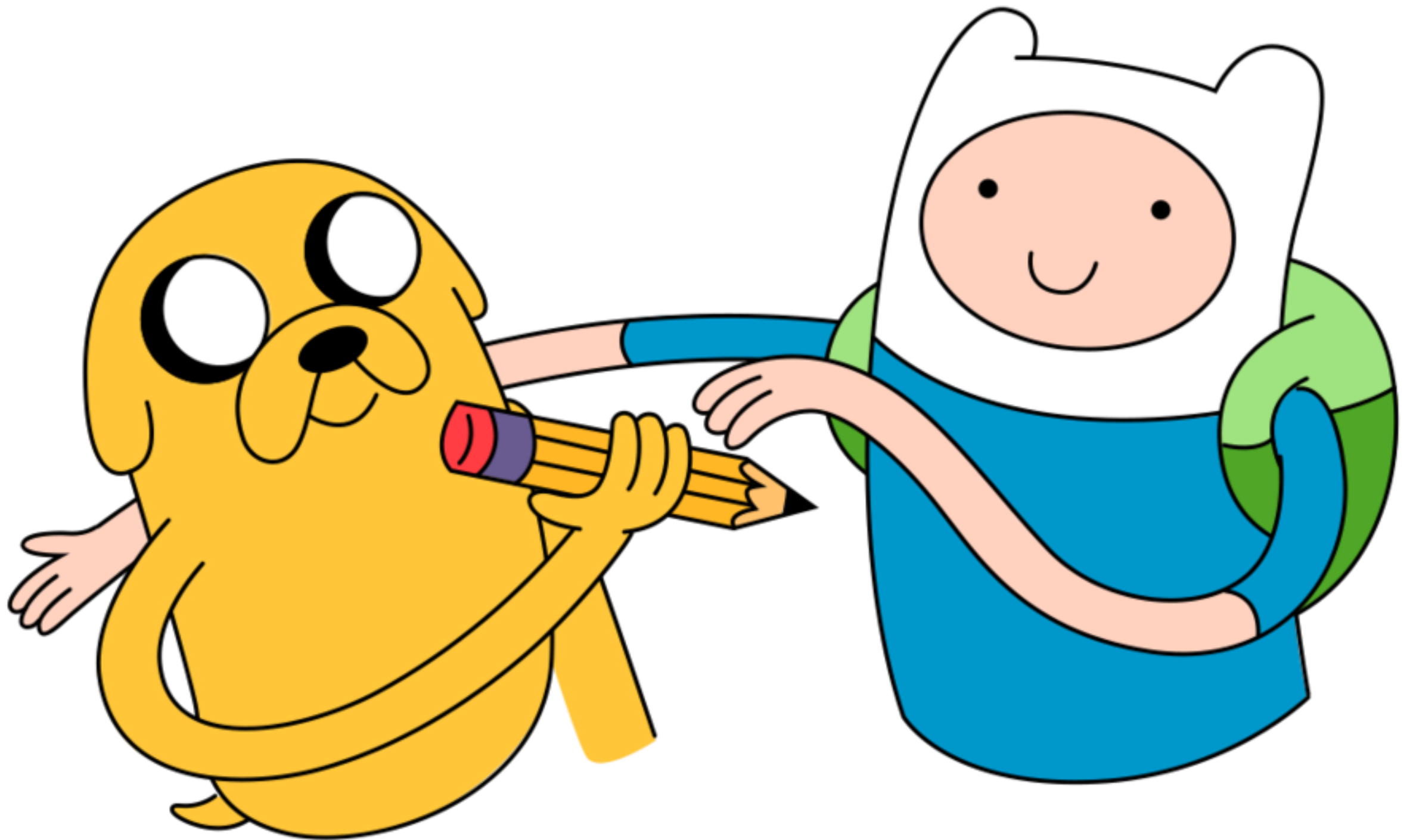
do
show
tell
read
watch
help
write up



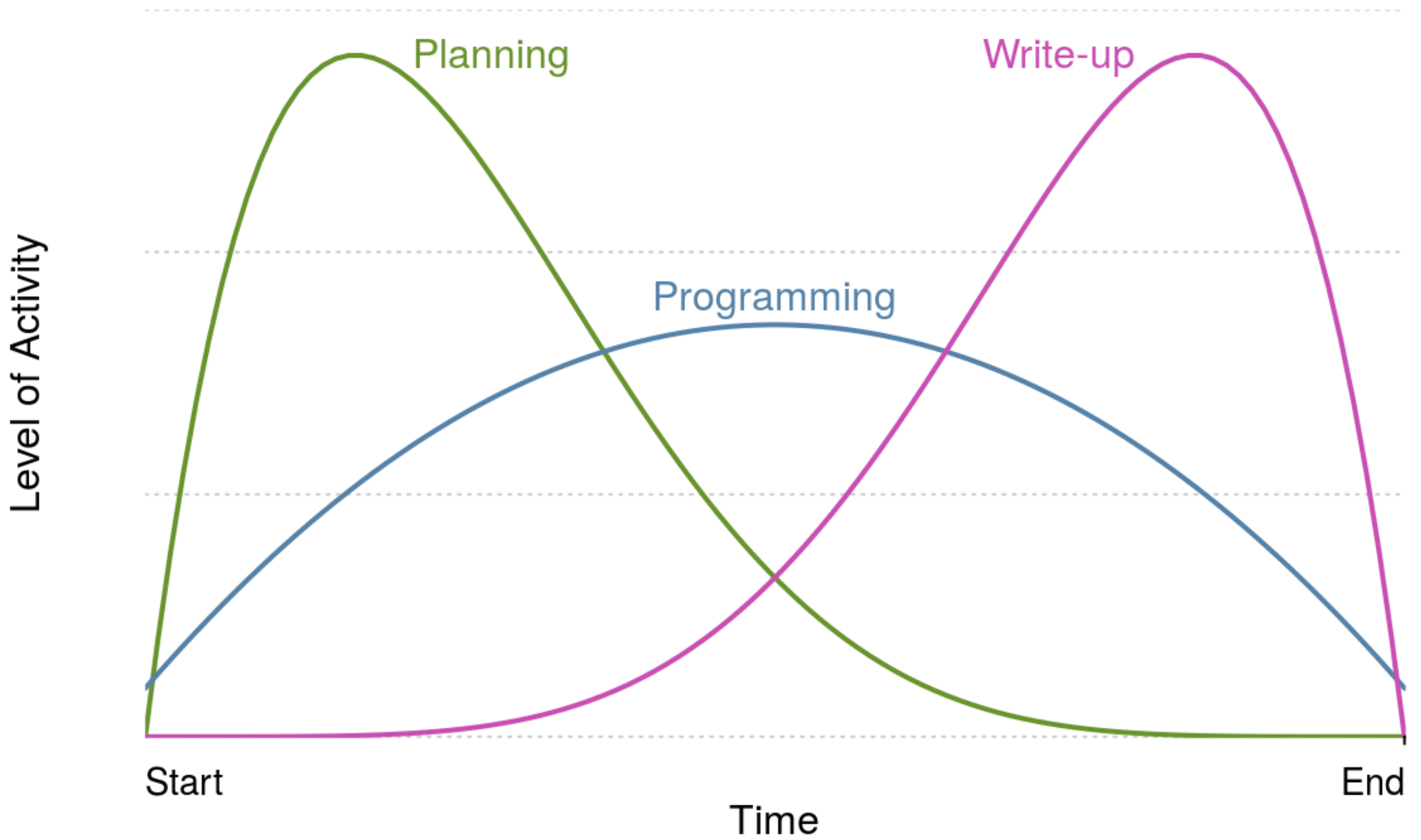


vignettes
books
stackoverflow
Journal of Statistical Software

efficient **planning**
of statistics before
code important



chunk your work



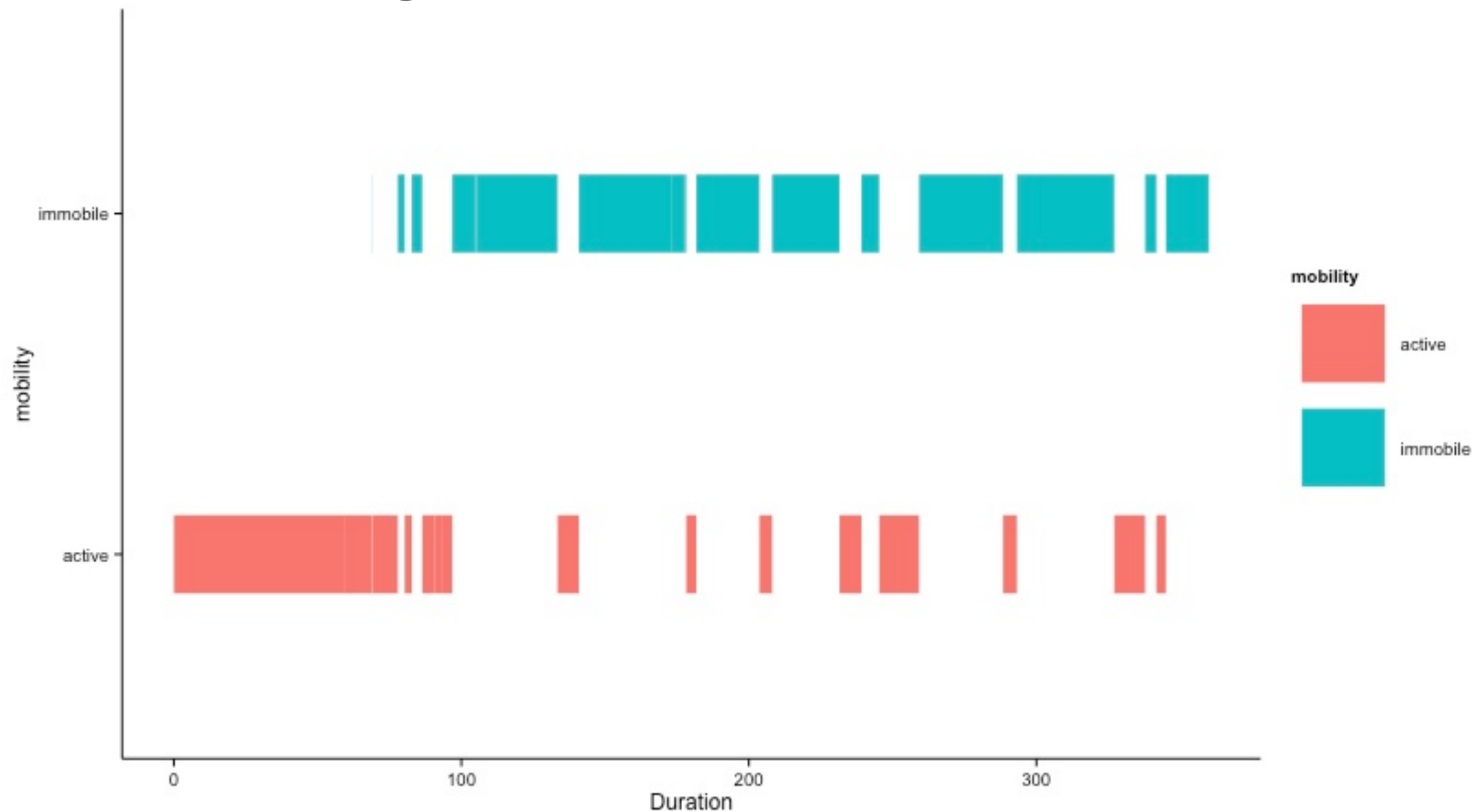
SMART workflows



specific
measurable
attainable
realistic
time-bound

efficient **planning directly** in R

plan
plotrix
DiagrammeR



Gantt charts

use packages: innovative & sometimes efficient

Adventure Time



Princesses



**Bee
Princess**



**Breakfast
Princess**



**Cotton Candy
Princess**



**Embryo
Princess**



**Emerald
Princess**



**Engagement Ring
Princess**



**Ghost
Princess**



**Hot Dog
Princess**



**Jungle
Princess**



**Lumpy Space
Princess**



**Mini
Princess**



**Muscle
Princess**



**Old Lady
Princess**



**Orange
Princess**



**Peanut
Princess**



**Princess
Beautiful**



**Princess
Bubblegum**



**Princess Princess
Princess**



**Purple
Princess**



**Turtle
Princess**



**Raggedy
Princess**



**Skeleton
Princess**



**Slime
Princess**



**Space Angel
Princess**



**Toast
Princess**



**Wildberry
Princess**

package selection

actively developed?

well documented?

well used?

efficient **set up** tips



monitor resources
use GitHub
RStudio

test code (microbenchmark)

update R & packages (`update.packages(ask = FALSE)`)

no ask-to-save & no restore defaults

directory management

efficient data **importing**



```
rio::import  
readr
```

```
.Rds
```

```
tip  
url <- ""  
df=read.csv(url)
```

efficient data **handling**



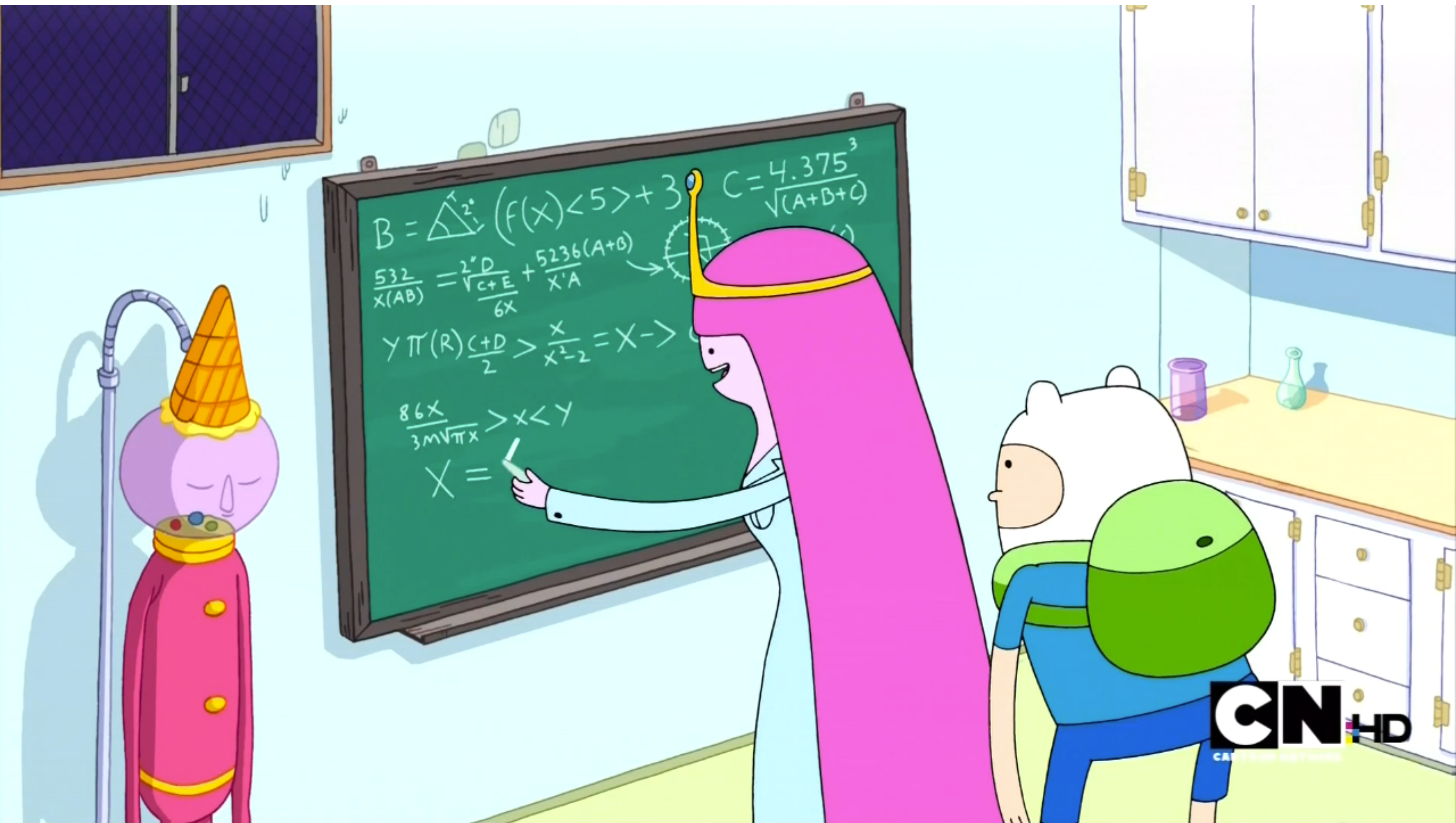
dplyr: data_grid
dplyr: drop_na
dplyr: tally
check class
%>%
tibbles
resample_bootstrap()

efficient **coding**



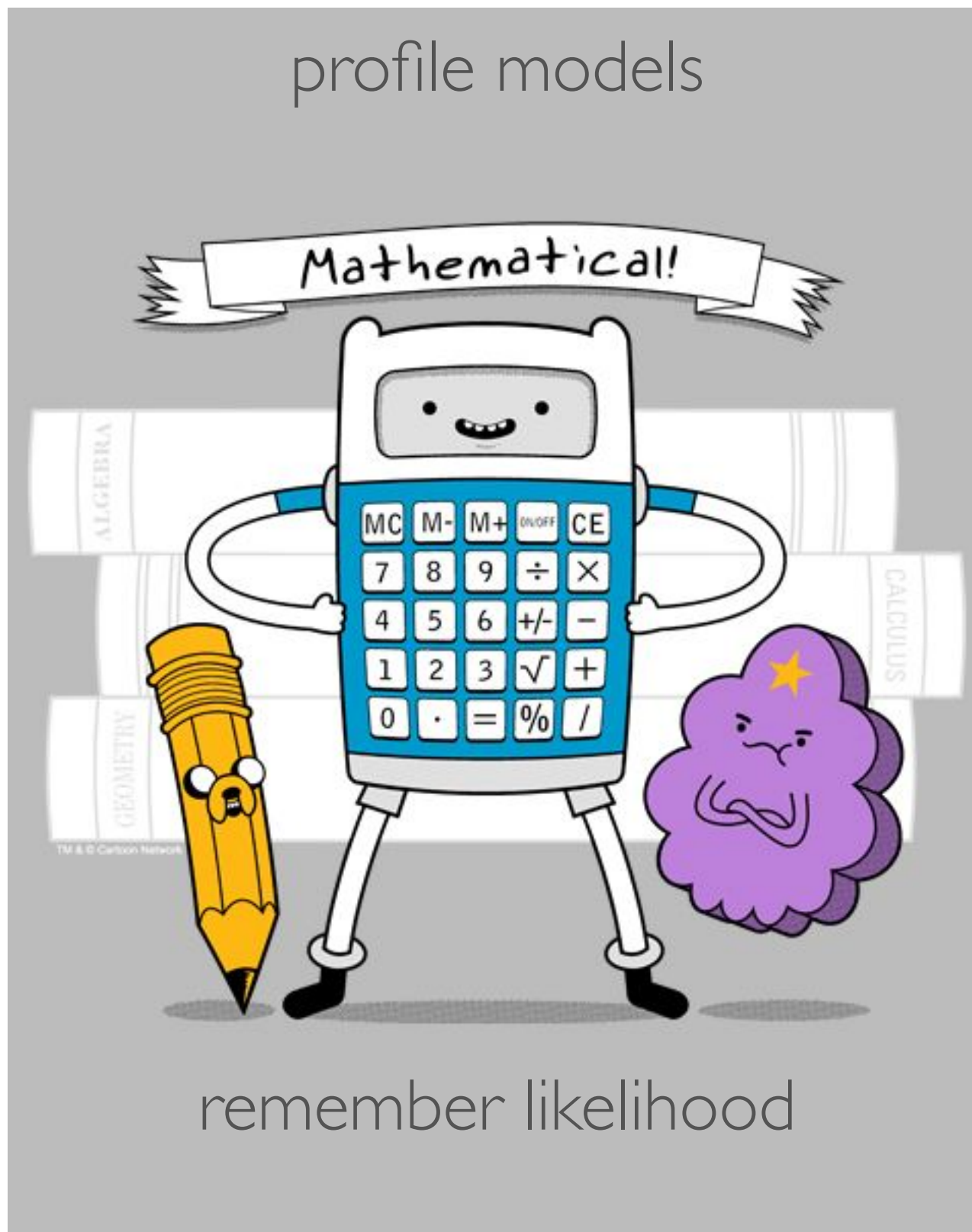
access underlying routines as quickly as possible
fewer functions is efficient
vectorize code: functions that work with all-length vectors

efficient **statistics**



modelr::model_matrix stats::glm() mgcv::gam()
glmnet::glmnet() MASS::rlm()

profile models

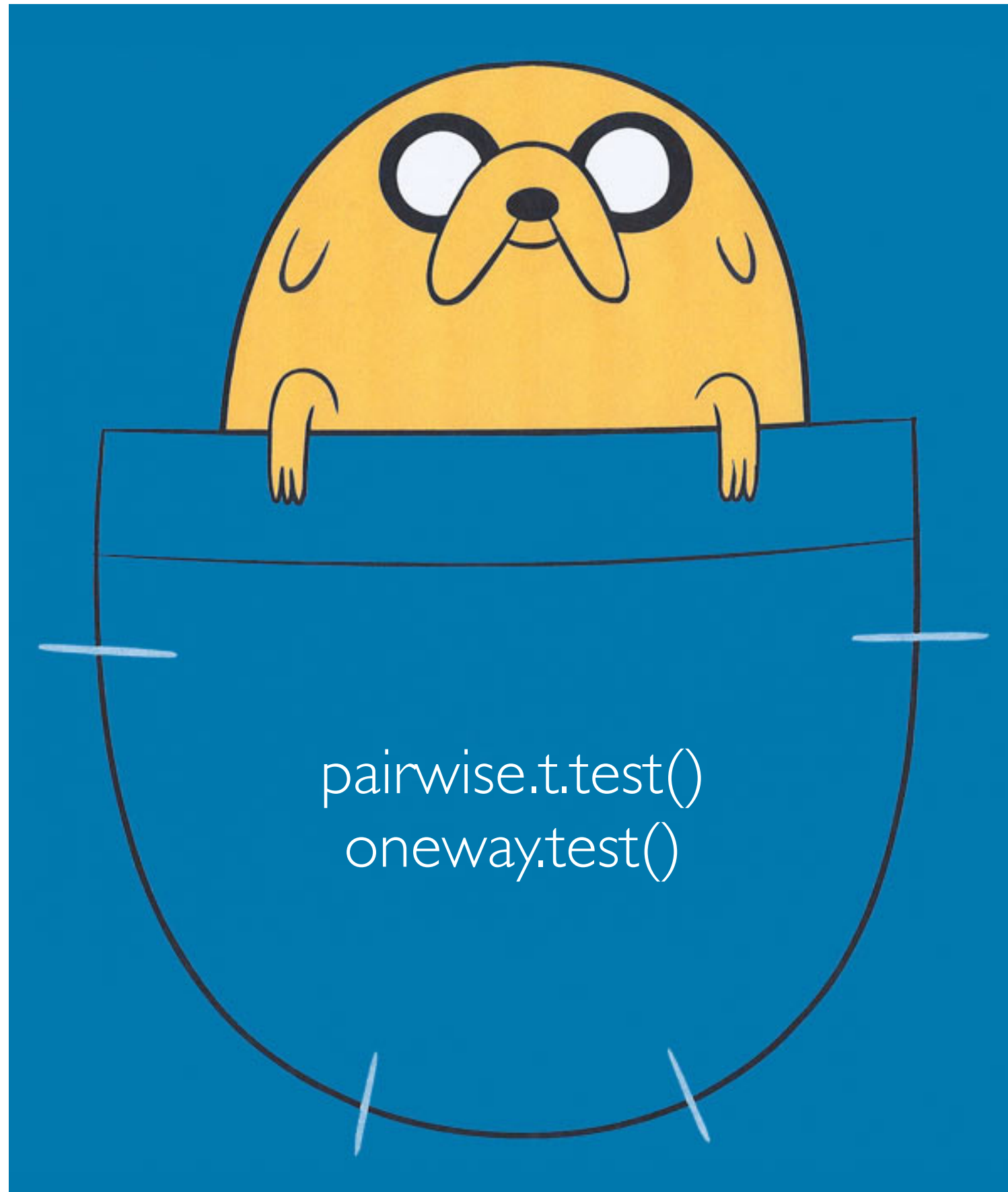


remember likelihood

data_grid and mine models



minimize assumptions





Efficient R Programming

A PRACTICAL GUIDE TO SMARTER PROGRAMMING

Colin Gillespie & Robin Lovelace