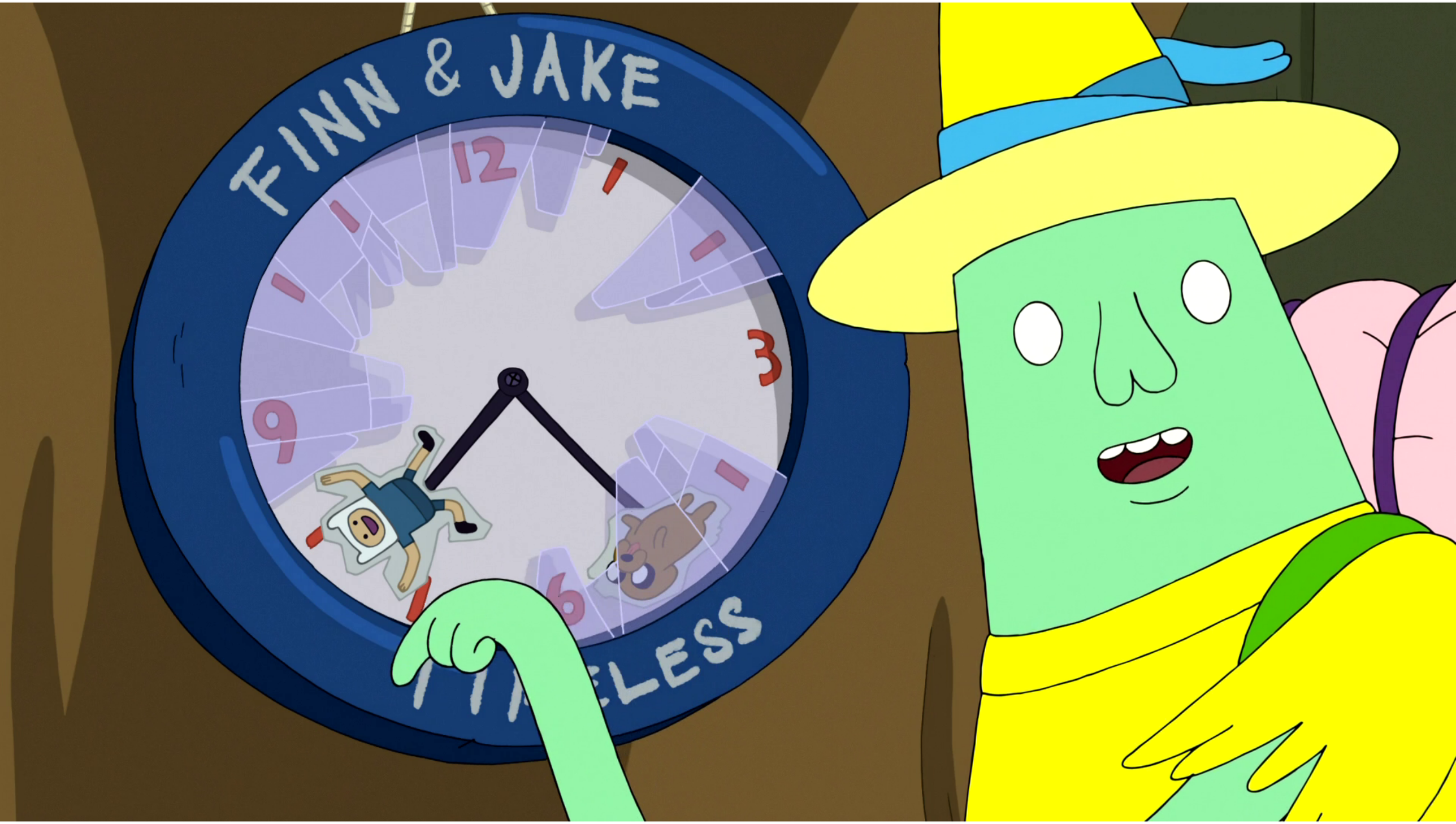
efficient statistics



@cjlortie

assumption: many domain scientists use R but are not programmers

R is the means to an end

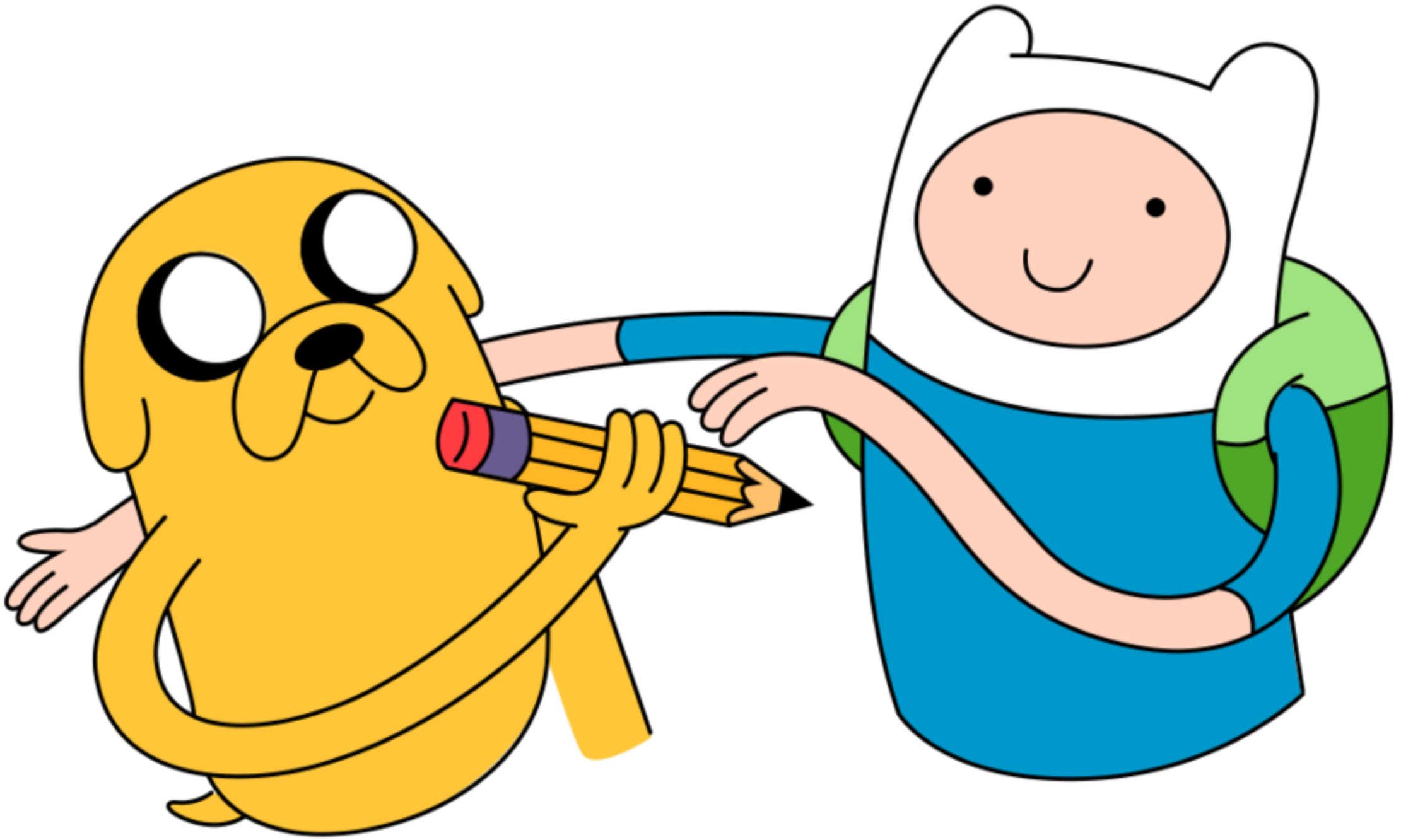
limitation: programming is a separate discipline

efficient **learning**
to code important

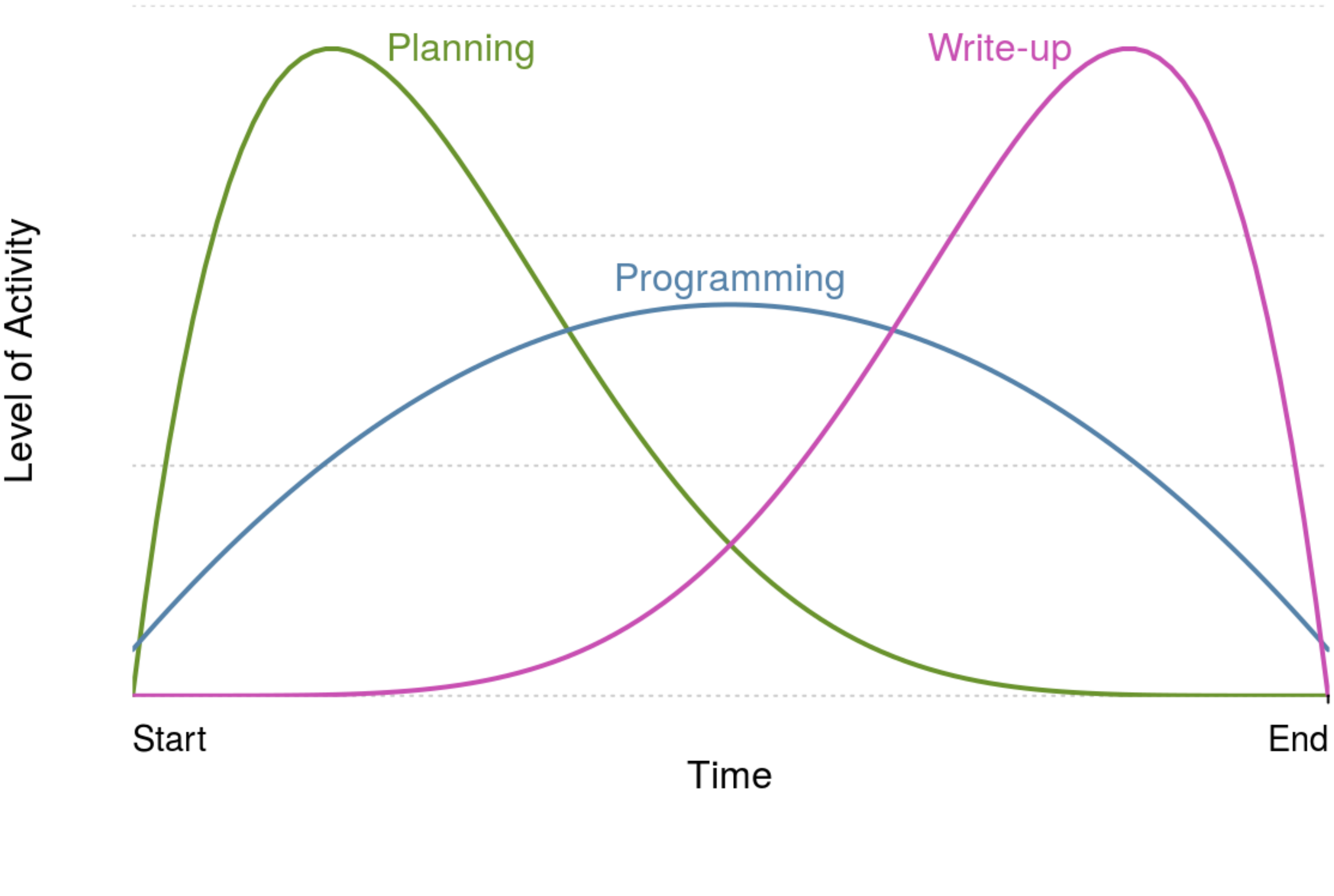
do
show
tell
read
watch
help
write up

vignettes
books
stackoverflow
Journal of Statistical Software

efficient **planning**
of statistics before
code important



chunk your work

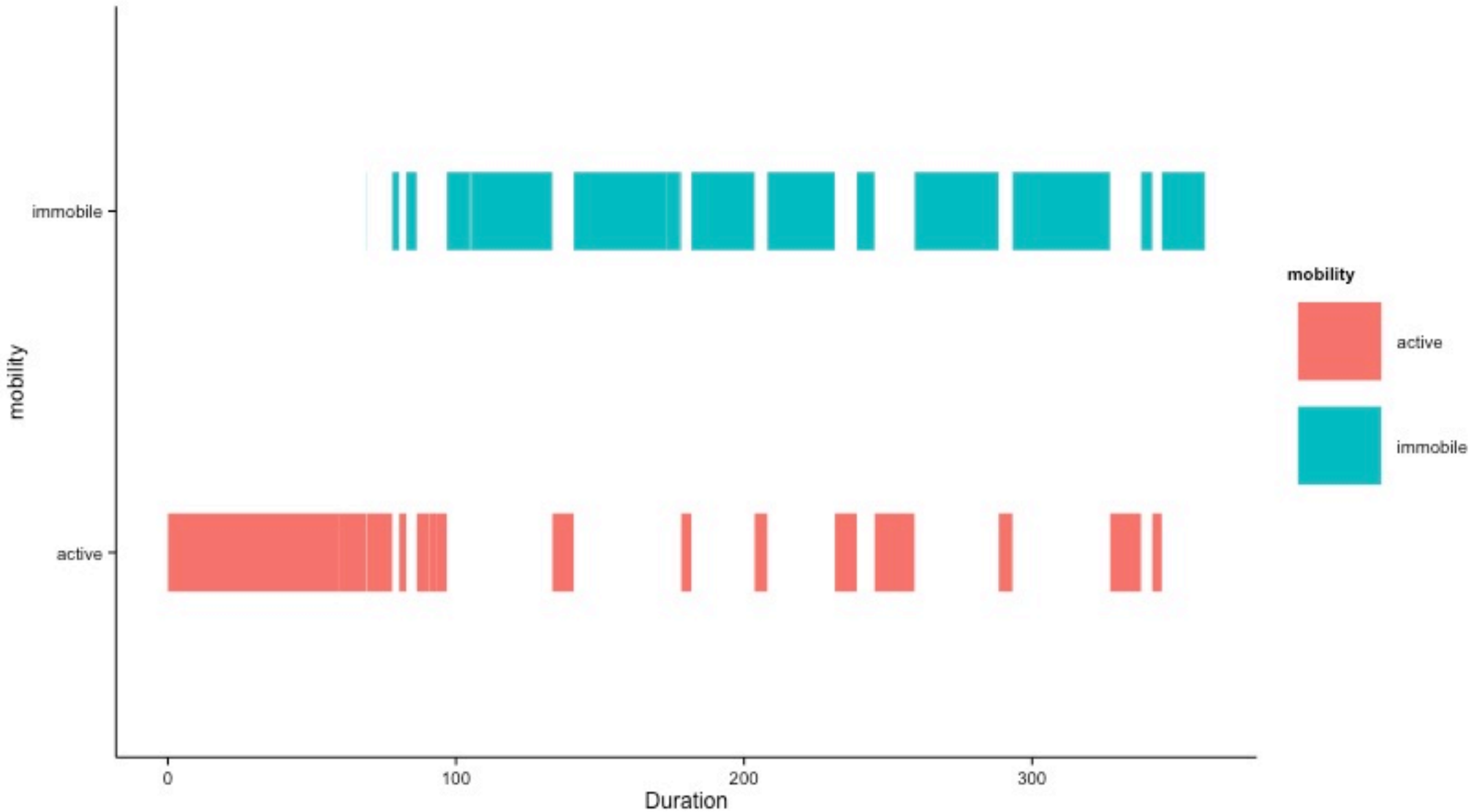


SMART workflows

Specific
measurable
attainable
realistic
time-bound

efficient **planning directly** in R

plan
plotrix
DiagrammeR



Gantt charts

use packages: innovative & sometimes efficient

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)

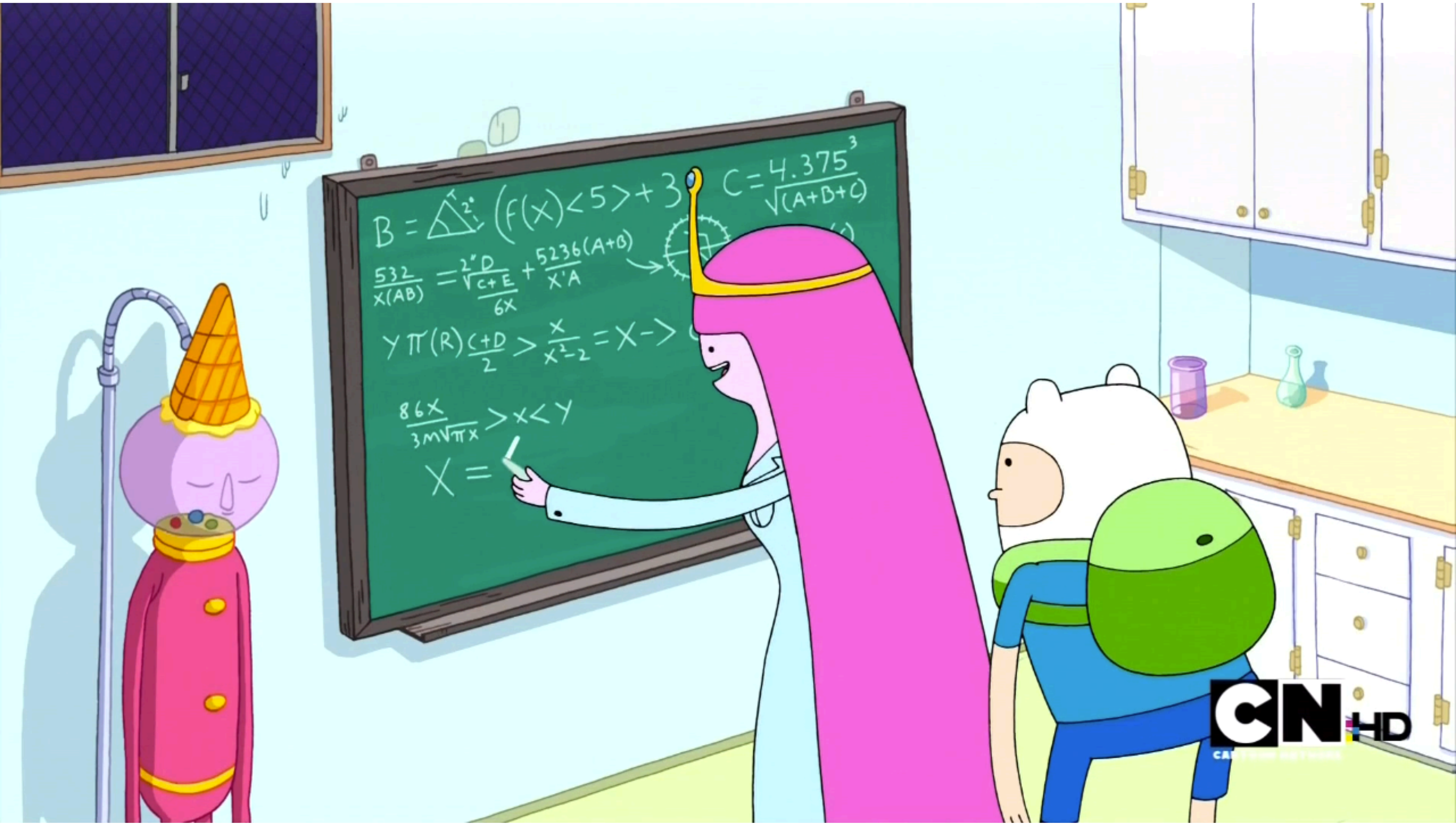
efficient data **handling**

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

efficient **coding**

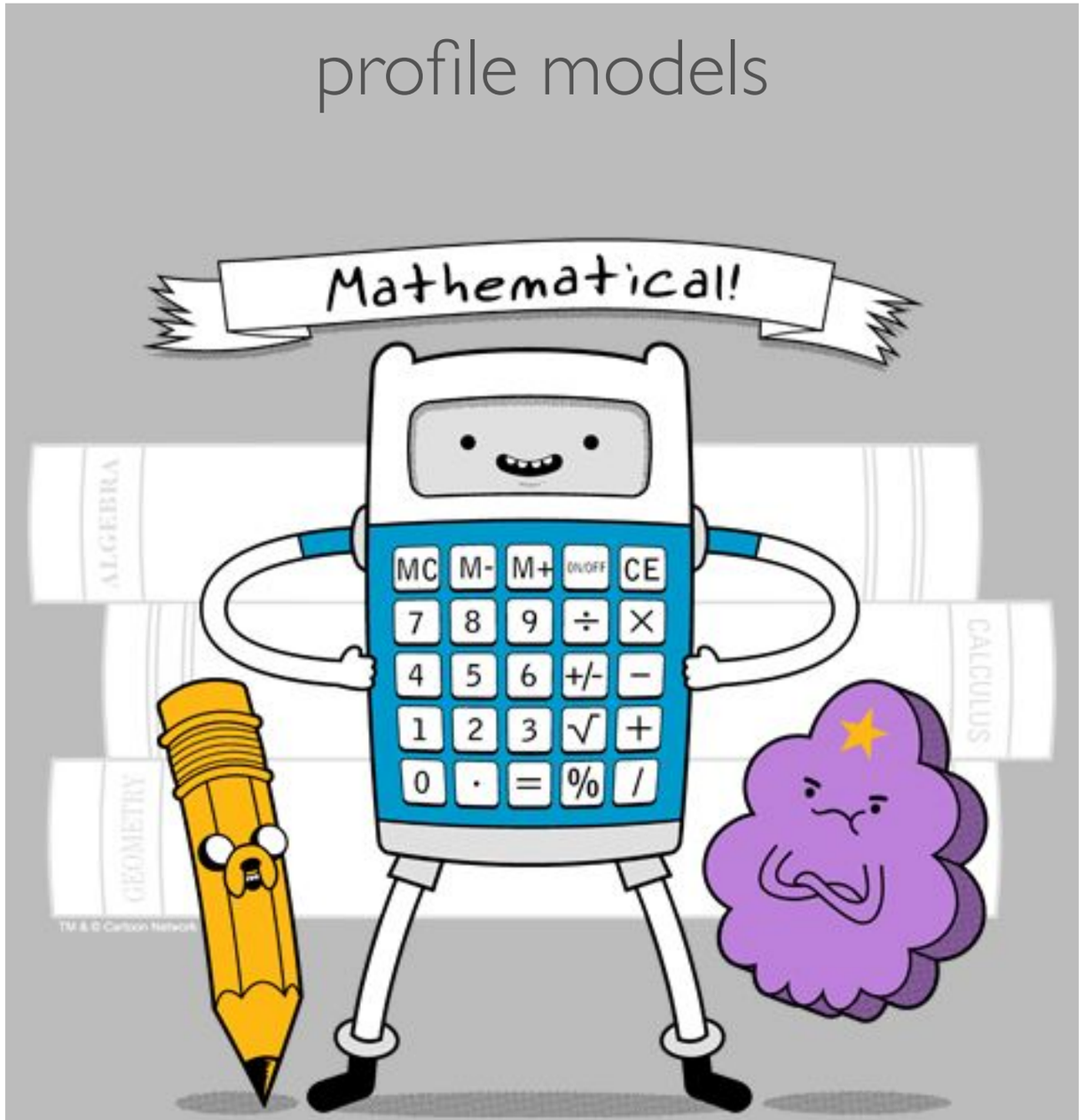
access underlying routines as quickly as possible
fewer functions is efficient
vectorize code: functions that work with all-length vectors

efficient **statistics**



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

profile models



remember likelihood

data_grid and mine models

minimize assumptions

O'REILLY®

Efficient R Programming

A PRACTICAL GUIDE TO SMARTER PROGRAMMING

Colin Gillespie & Robin Lovelace