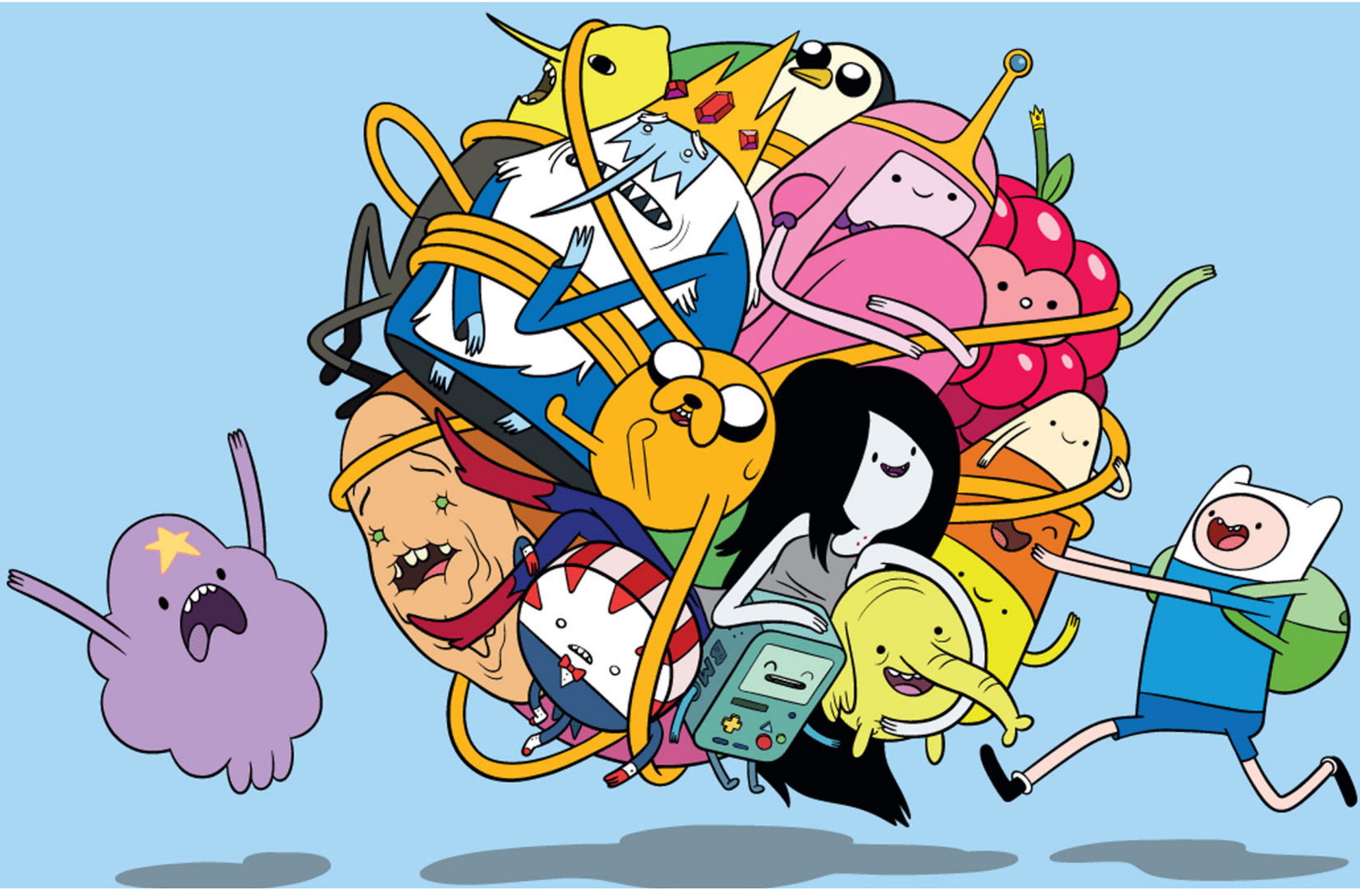
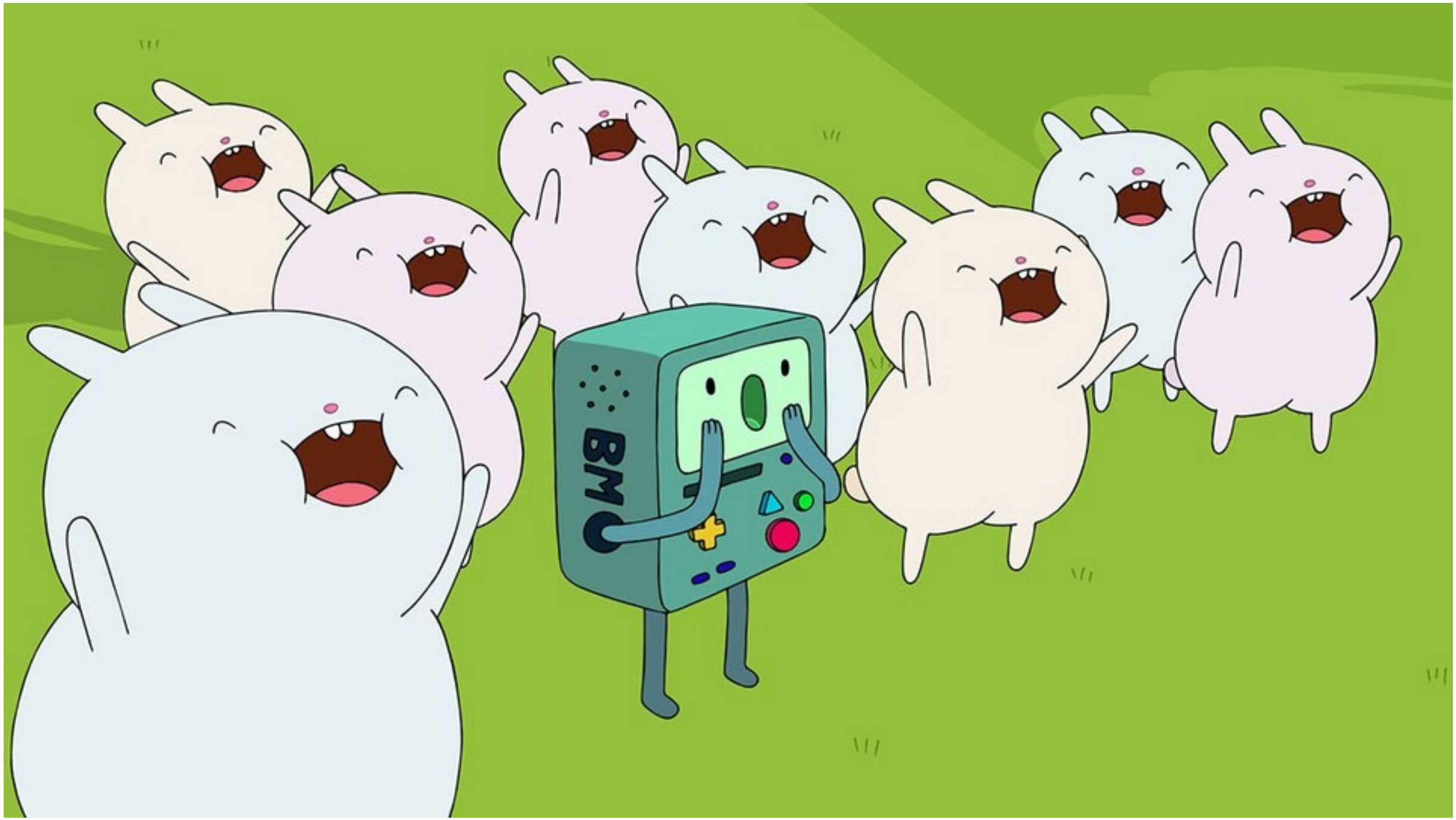
exploratory data analysis & **models** in R



@cjlortie

untangle, describe, predict, share, advance understanding

read data/build frame
inspect data structure
visualize data
EDA
fit model



EDA != the statistical model

EDA

creative adventures with your data

EDA

structure
scope of inference
data quality

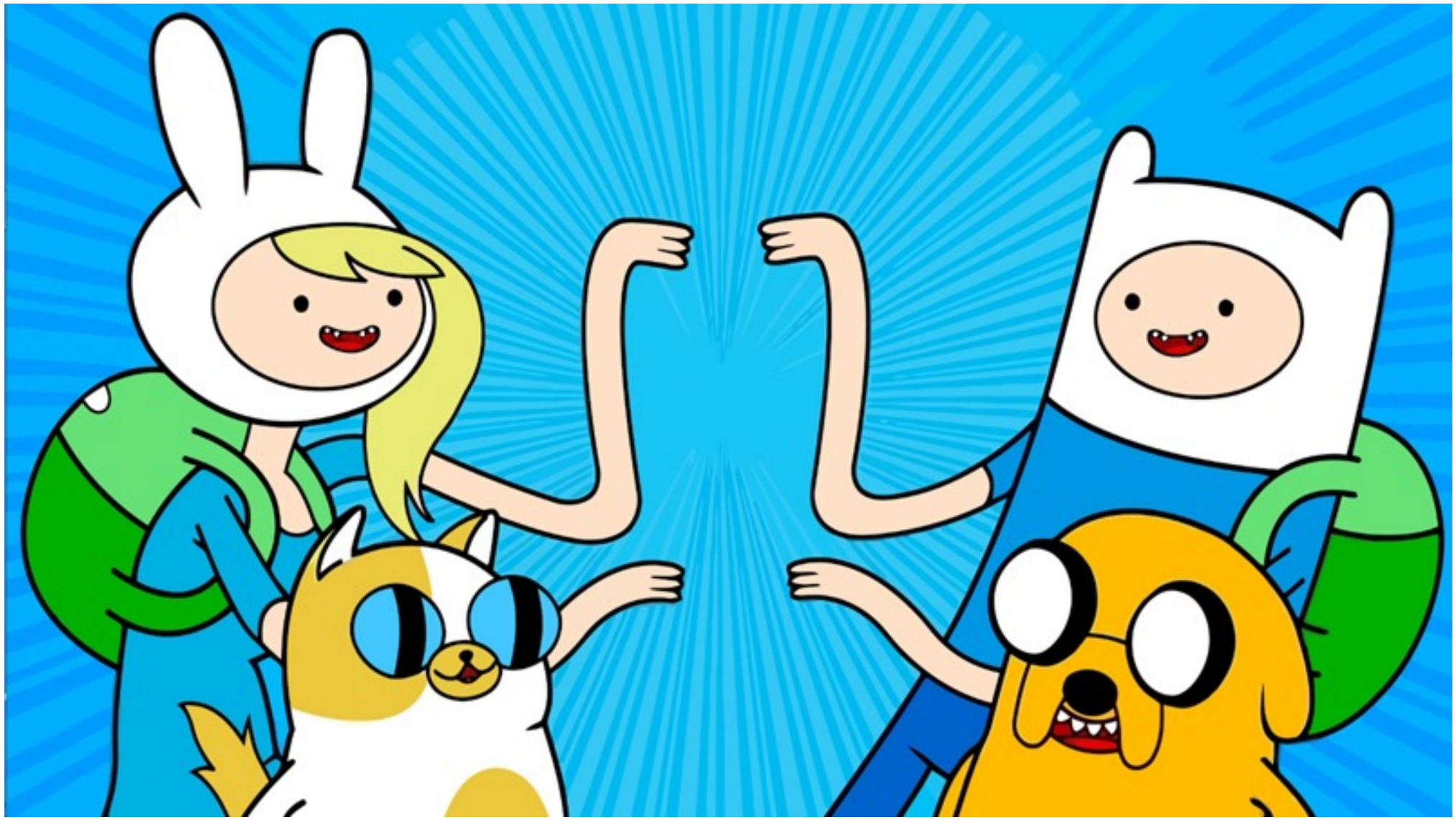
statistical model

an elegant, representative simplification of the patterns identified through data viz & EDA

model

convey structure
support evidence
estimate probability

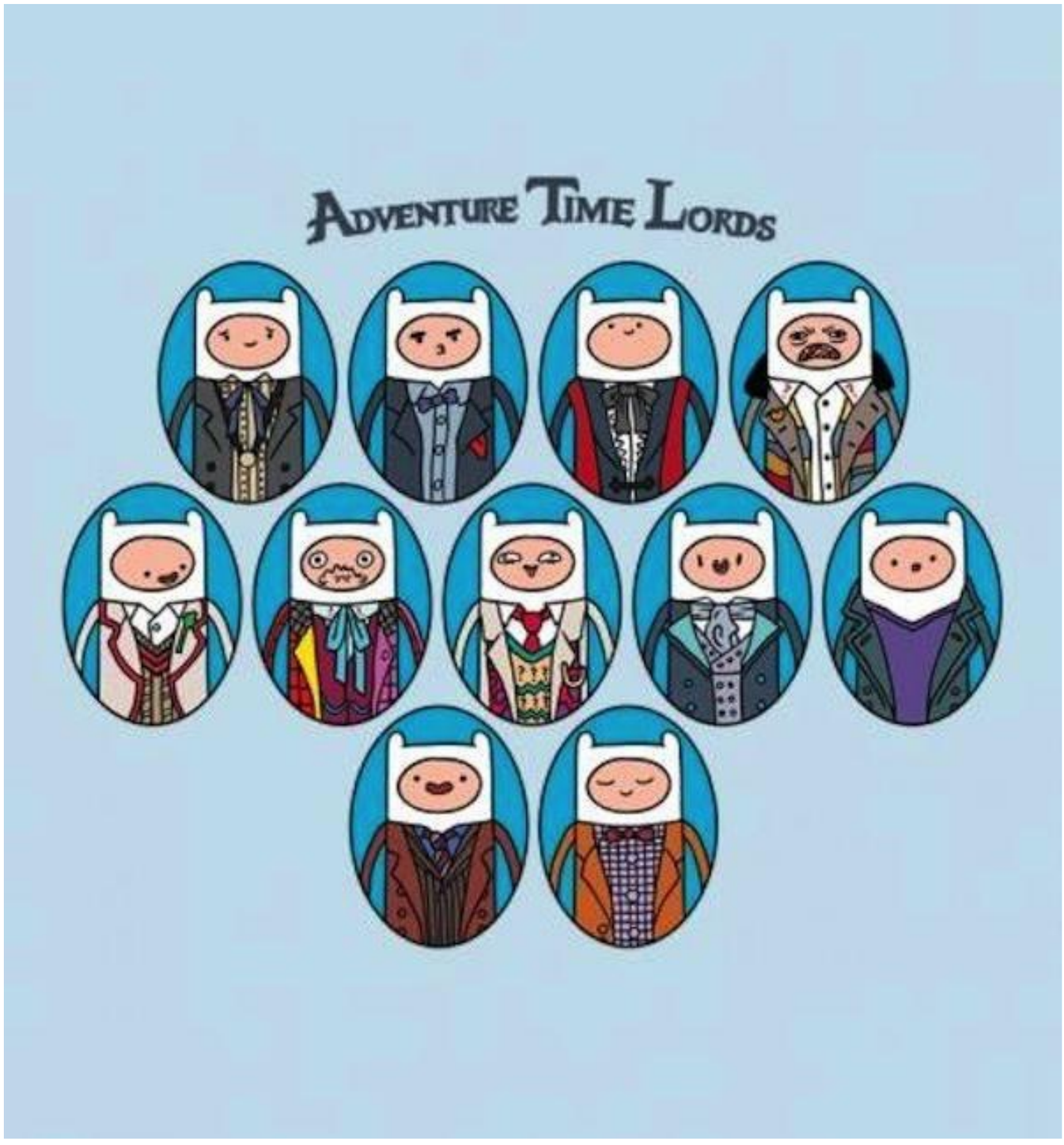
two kinds of statistical models



descriptive or predictive

time travel
backward or forward
but not both

effective model use is still **process** not product



effect sizes and coefficients

fit, inspect, mine

ZEZZ

17.10.2014

Symbol	Example	Meaning
+	+X	include this variable
−	−X	delete this variable
:	X:Z	include the interaction between these variables
*	X*Y	include these variables and the interactions between them
|	X|Z	conditioning: include x given z
^	(X + Z + W)^3	include these variables and all interactions up to three way
I	I(X*Z)	as is: include a new variable consisting of these variables multiplied
1	X - 1	intercept: delete the intercept (regress through the origin)

Dalgaard 2008

t.test to GLMM

summary and coef

predict
make_prediction
add_predictions
add_residuals

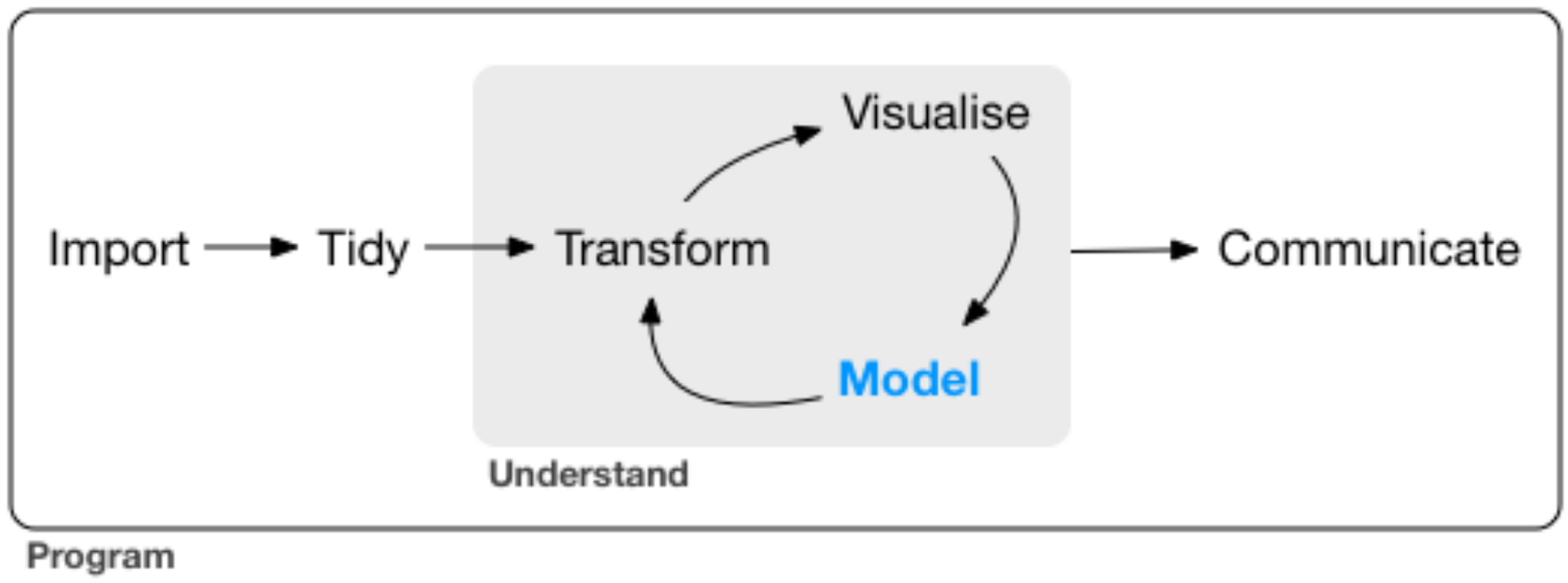
train & test

predictions are for new outcomes & naive data

bootstrapping

fixed or random effects

big picture of models is to advance understanding



'R for Data Science' convergence with R for statistics