

The Power of Great Datasets

Alex Cookson

 @alexcookson

 tacookson

Thank You Allison Horst for the amazing illustration ❤️

How do I learn how to
do [thing] in R?

How do I learn how to
do [thing] in R?

[sentiment analysis]

[data visualization]

[web scraping]

[forecasting]

[...]



What about

Great Datasets?

Great Datasets are

COOL

and

INTERESTING

Great Datasets are

"That's cool!"

COOL

Speaks to you on
a deeper level

Cool to me isn't
(necessarily)
cool to you

and

INTERESTING

Great Datasets are

COOL

and

Rouses your
curiosity

Prompts you to
ask questions

INTERESTING

"Just one
more graph"

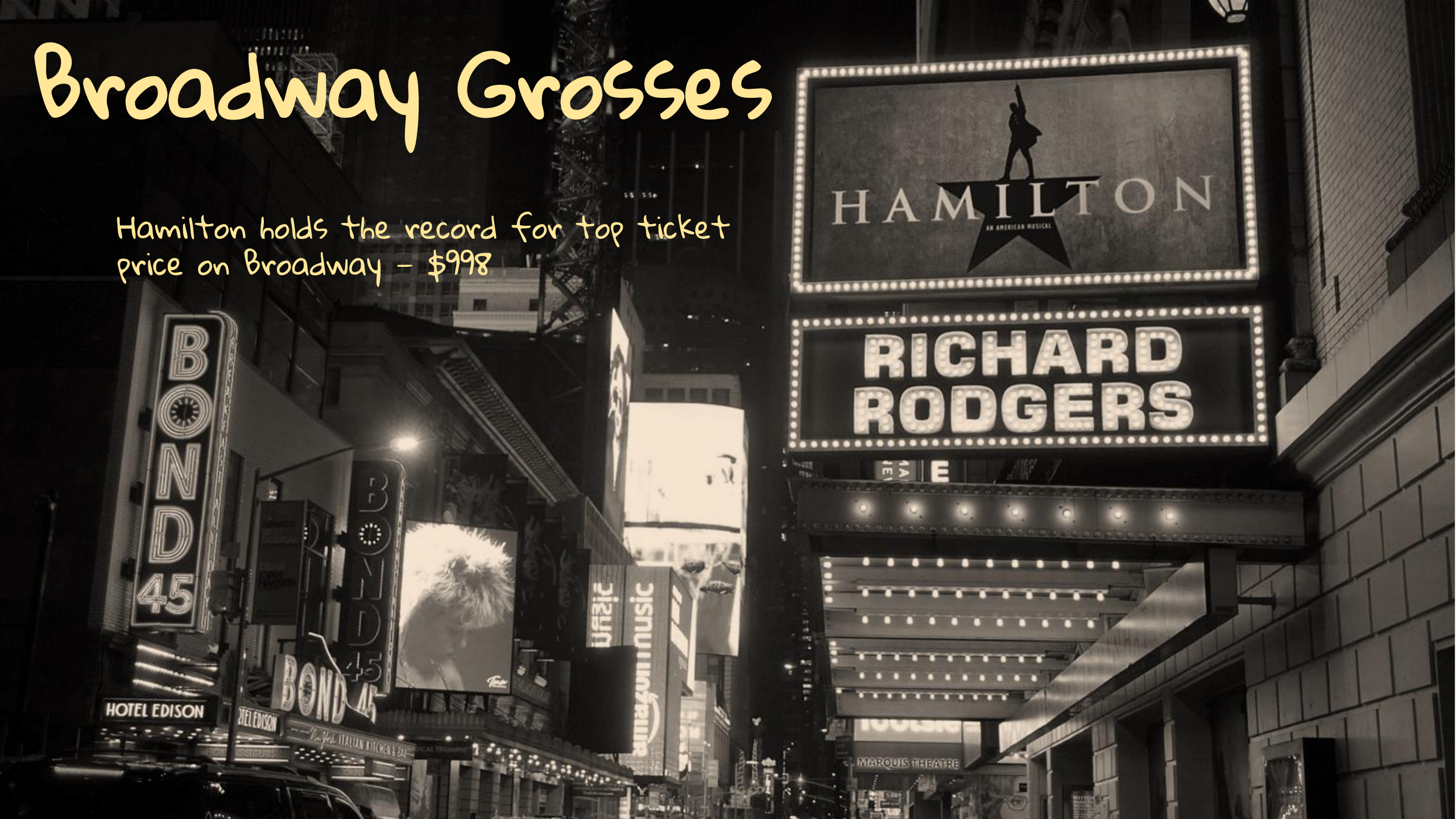
Duke Lemur Center



Ring-tailed Lemurs have an average weight of 2.2kg

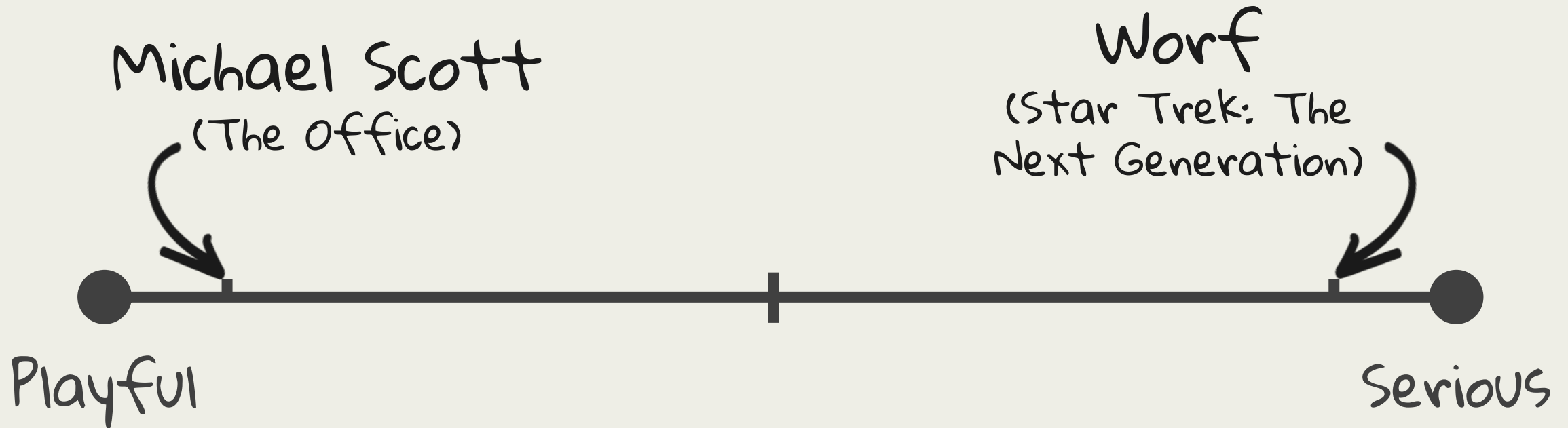
Broadway Grosses

Hamilton holds the record for top ticket price on Broadway – \$998



Fictional Character Personalities

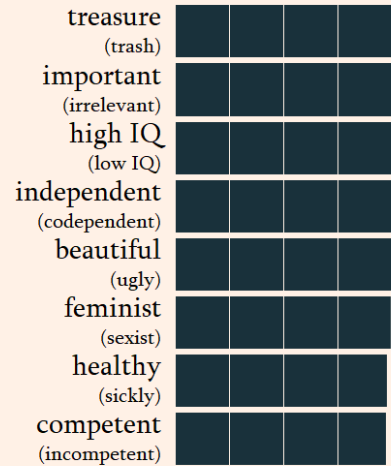
Traits of 800 characters across 250+ spectrums



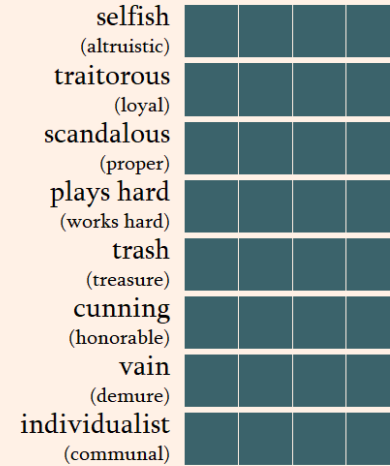
What are some of my favourite
characters' personalities?

Pride and Prejudice

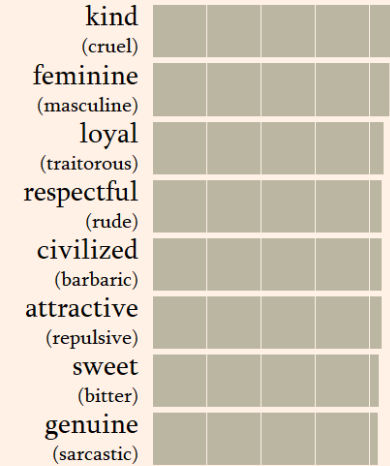
Elizabeth Bennet



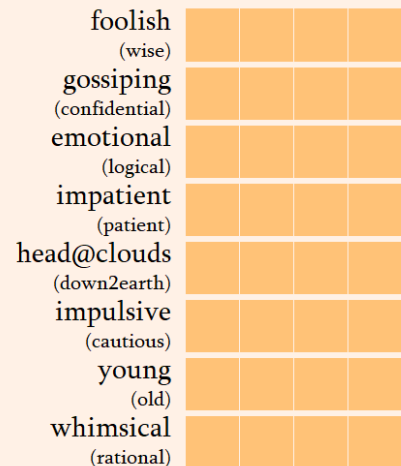
George Wickham



Jane Bennet

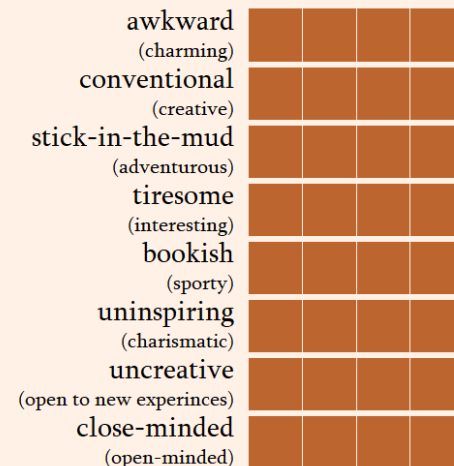


Lydia Bennet



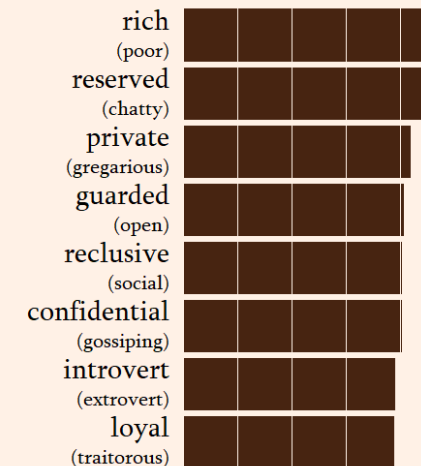
0 10 20 30 40 50

Mr. Collins



0 10 20 30 40 50

Mr. Darcy



0 10 20 30 40 50

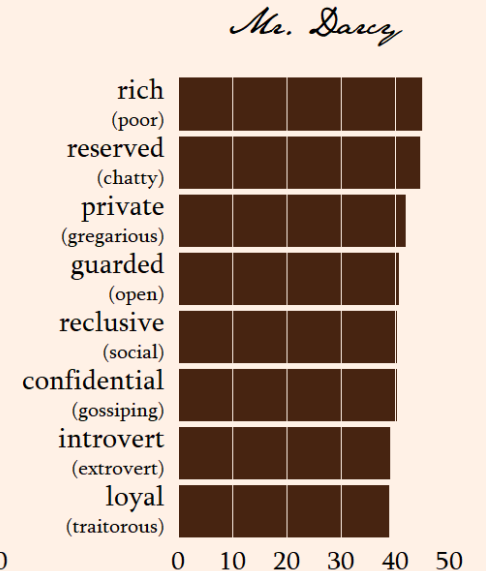
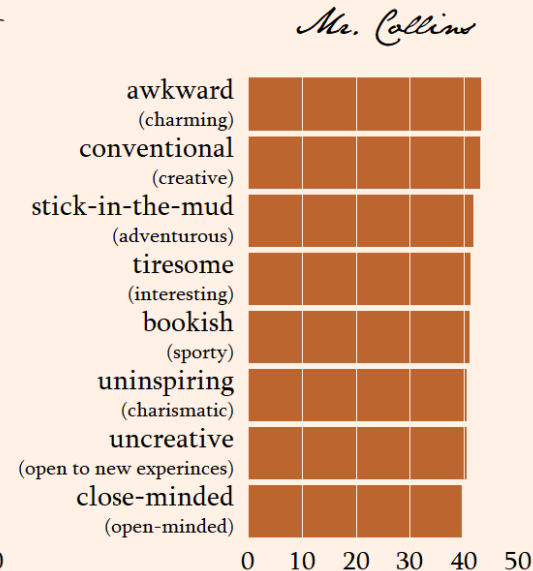
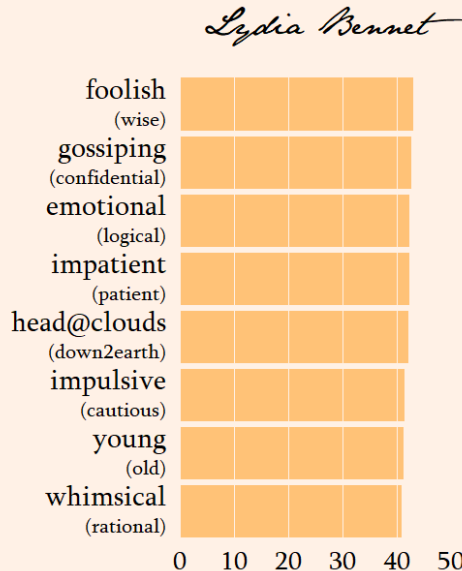
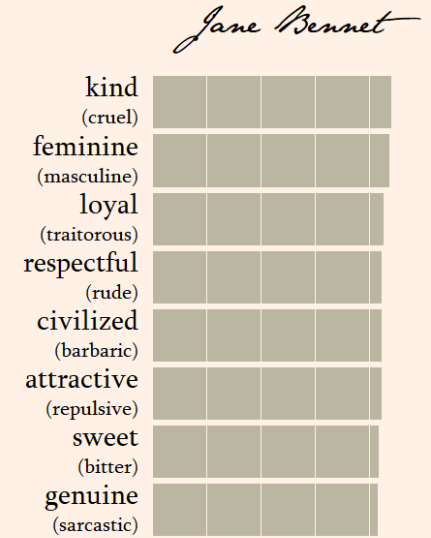
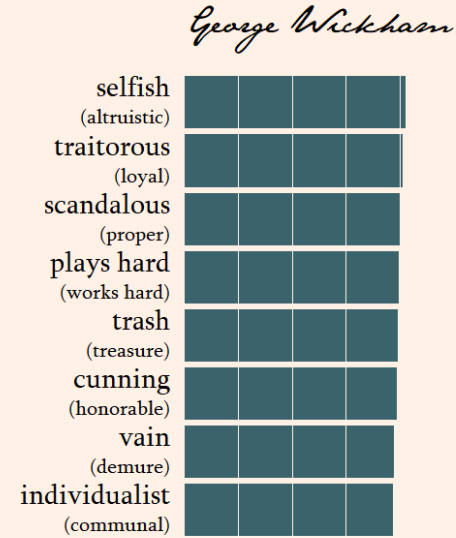
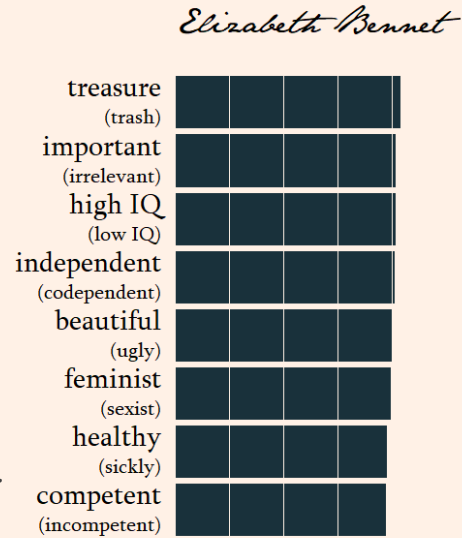
Use custom fonts

Pride and Prejudice

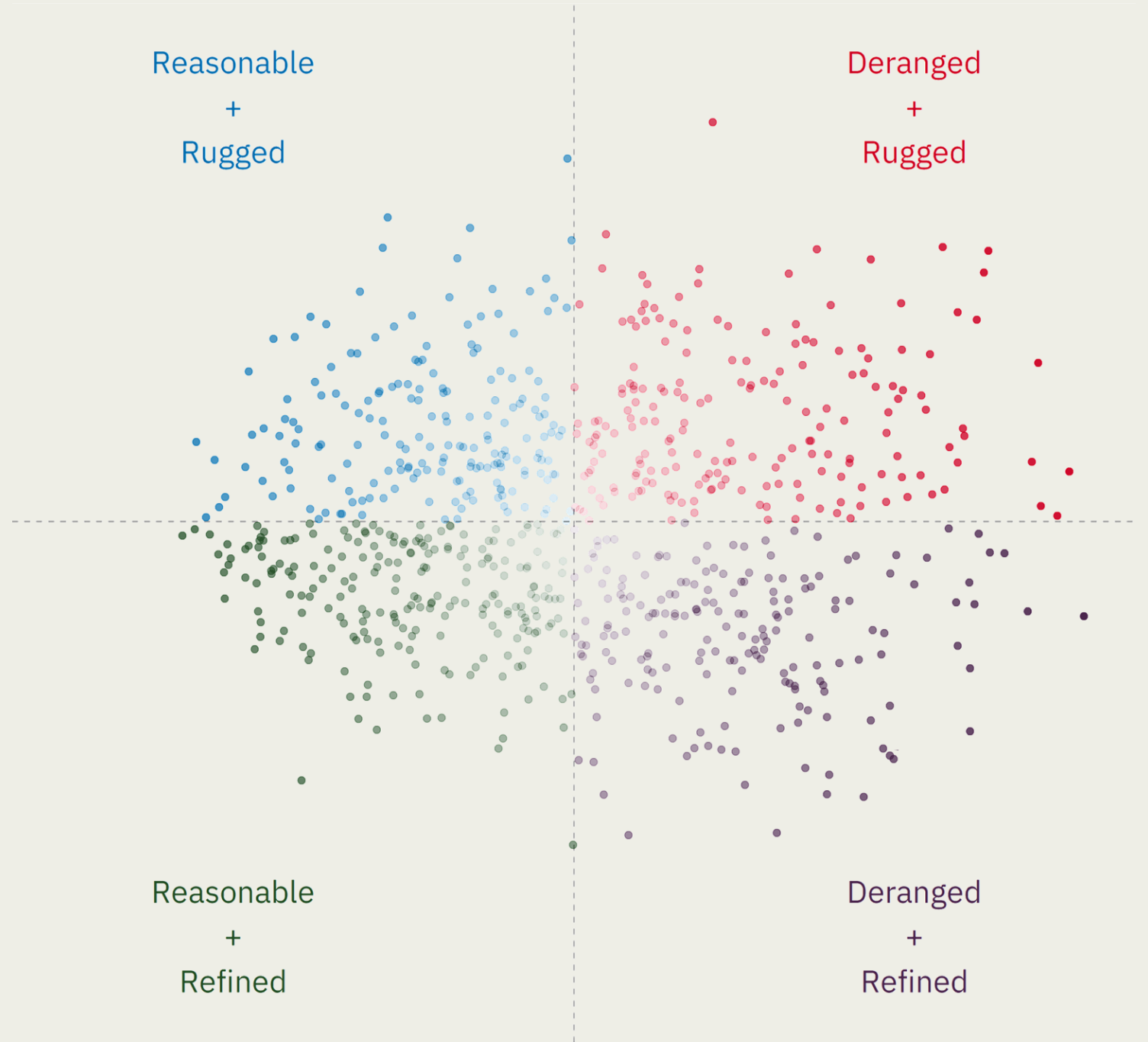
Tweak labels with
{glue} and {ggtext}

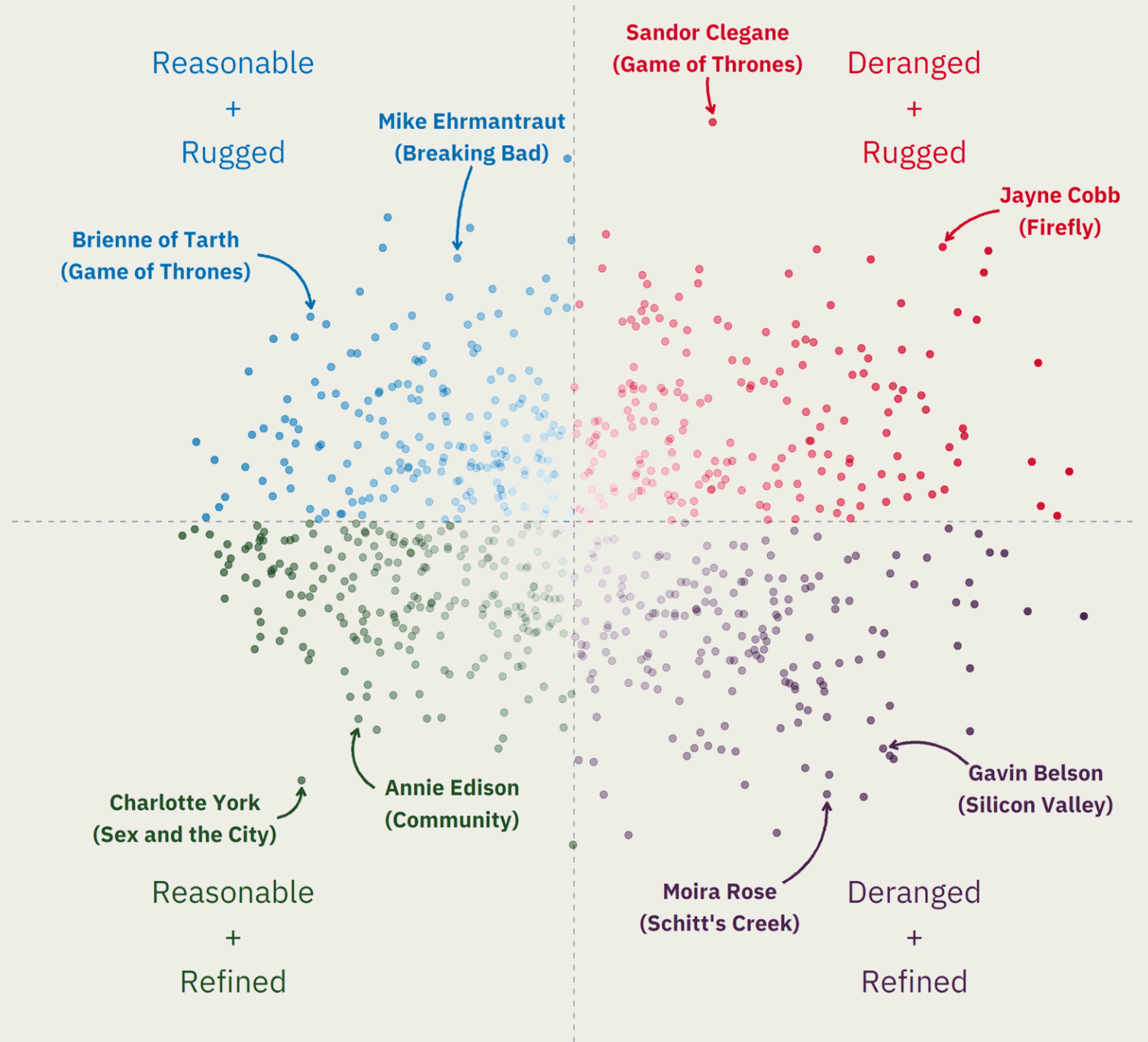
Choose an appropriate
colour palette

Adjust a LOT of
theme() elements



Which characters are most similar to one another?

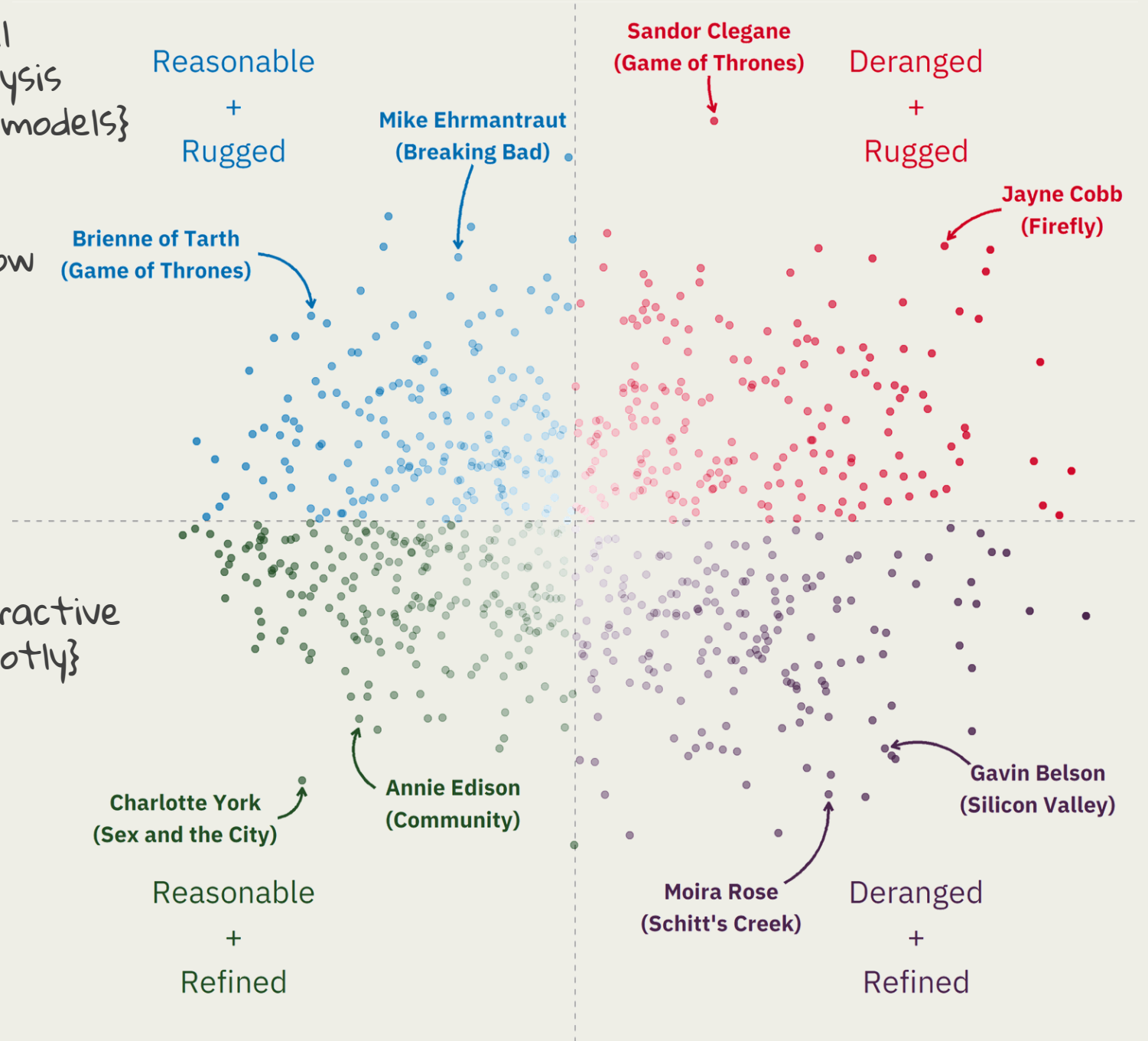




Conduct Principal
Component Analysis
(PCA) with {tidymodels}

Add text and arrow
annotations

Make an interactive
chart with {plotly}



I'll be honest...

I would **never** put this much
effort into analyzing
Titanic passengers





The Power of Great Datasets

Alex Cookson

 @alexcookson

 tacookson

Thank You Allison Horst for the amazing illustration ❤️