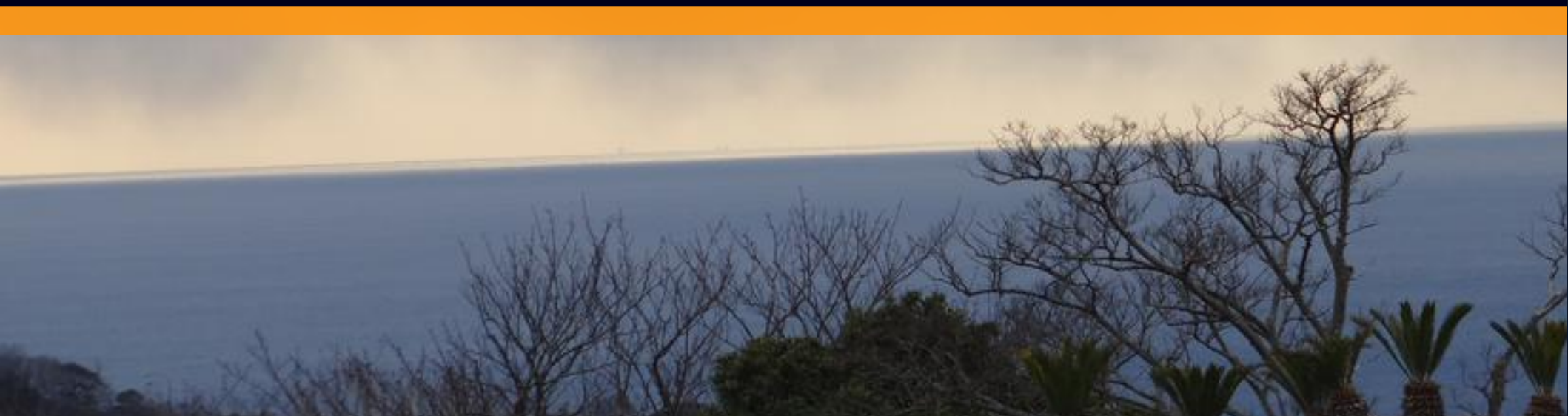


Introduction to Data Science: Course Project

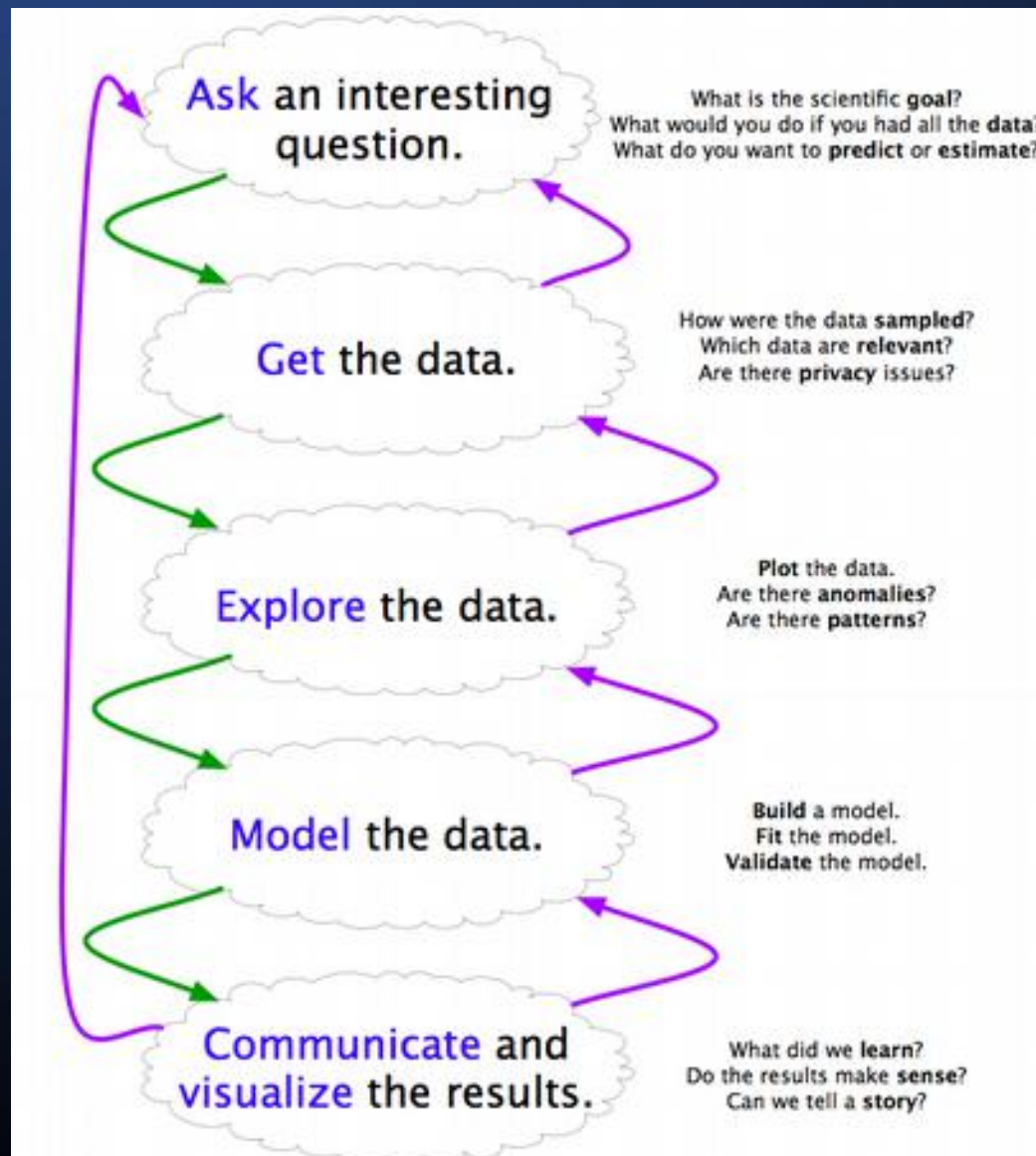
Gerard de Melo

<http://gerard.demelo.org>

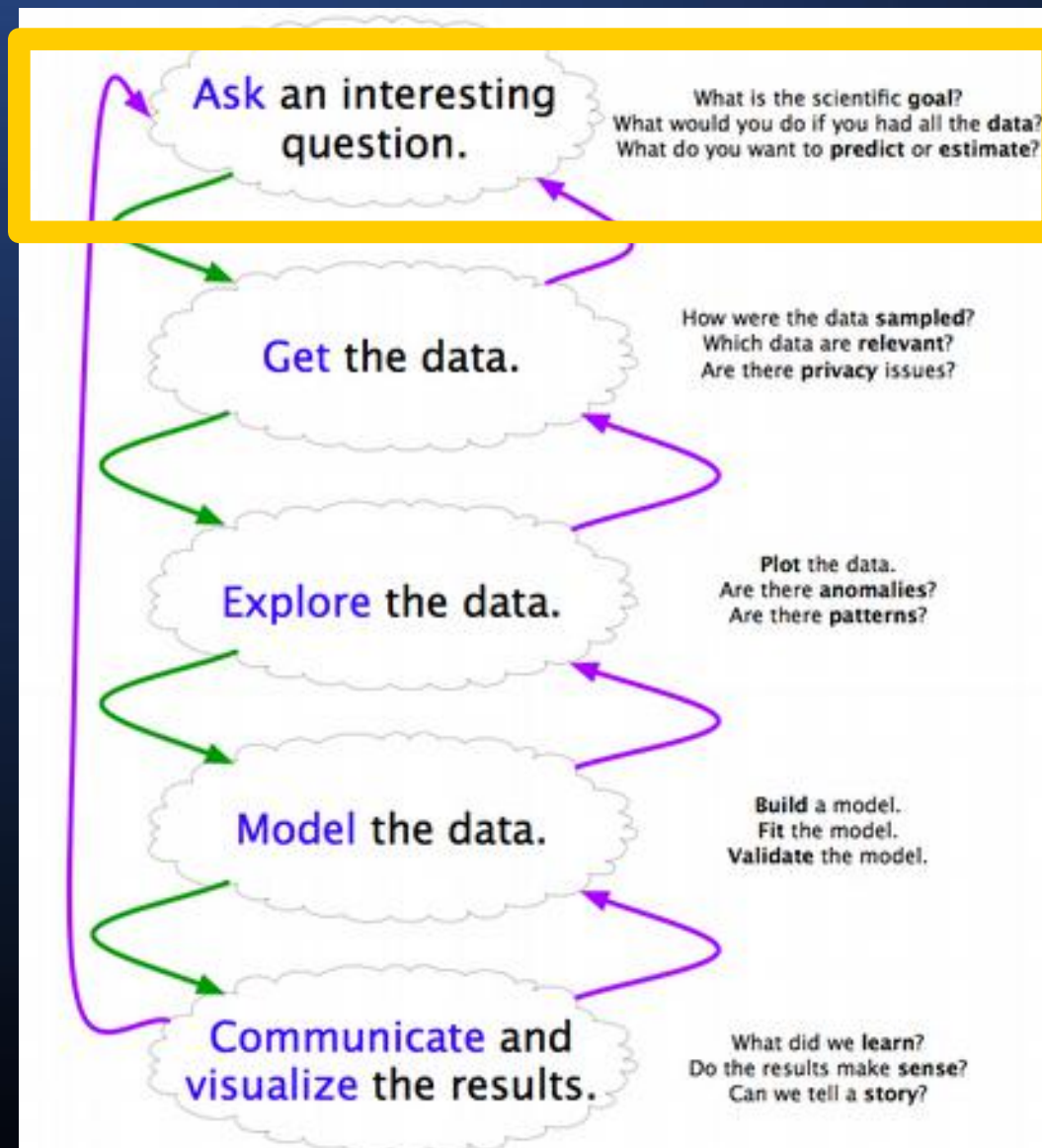
Rutgers University



Course Project



Course Project



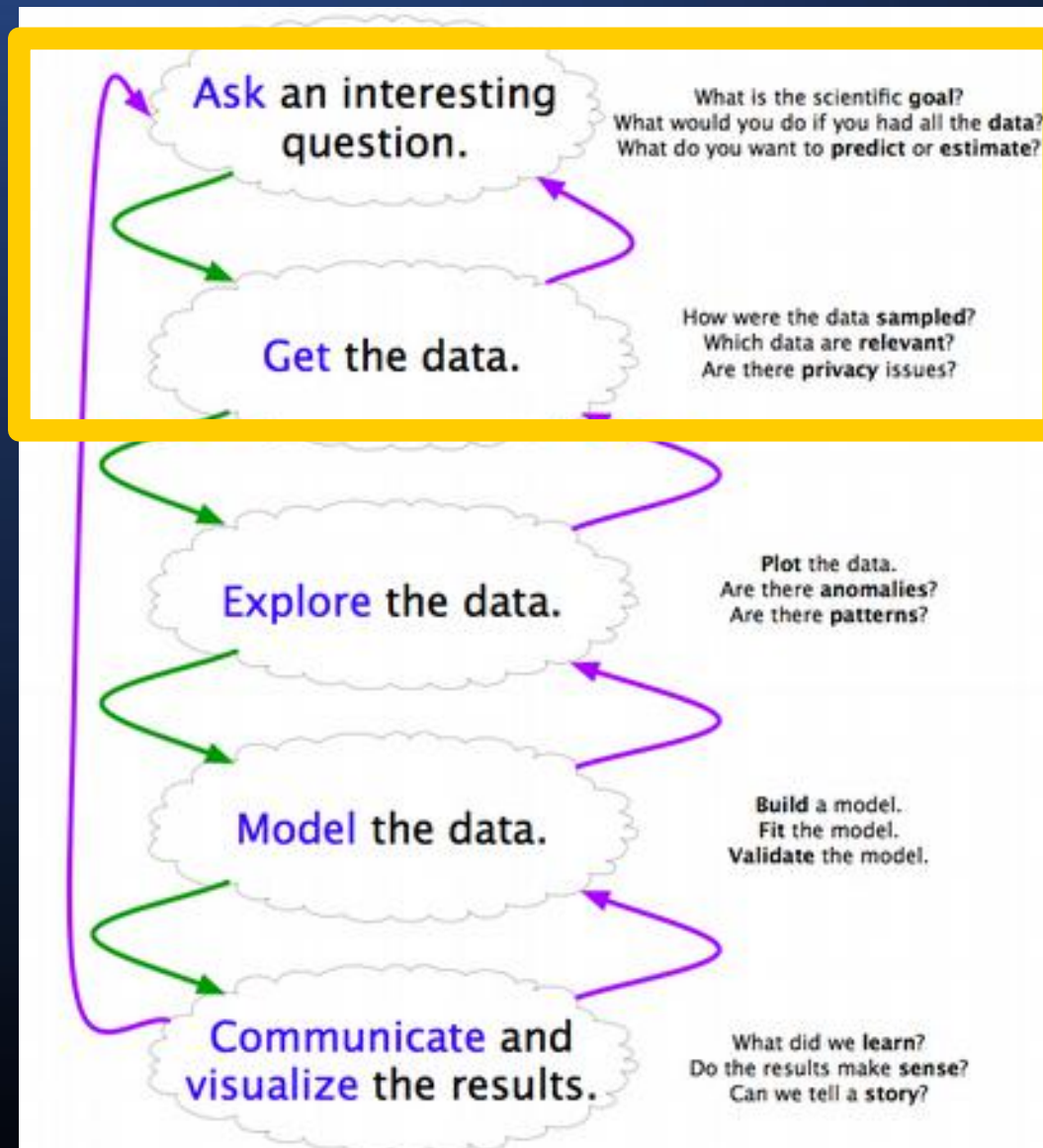
Course Project: Topic



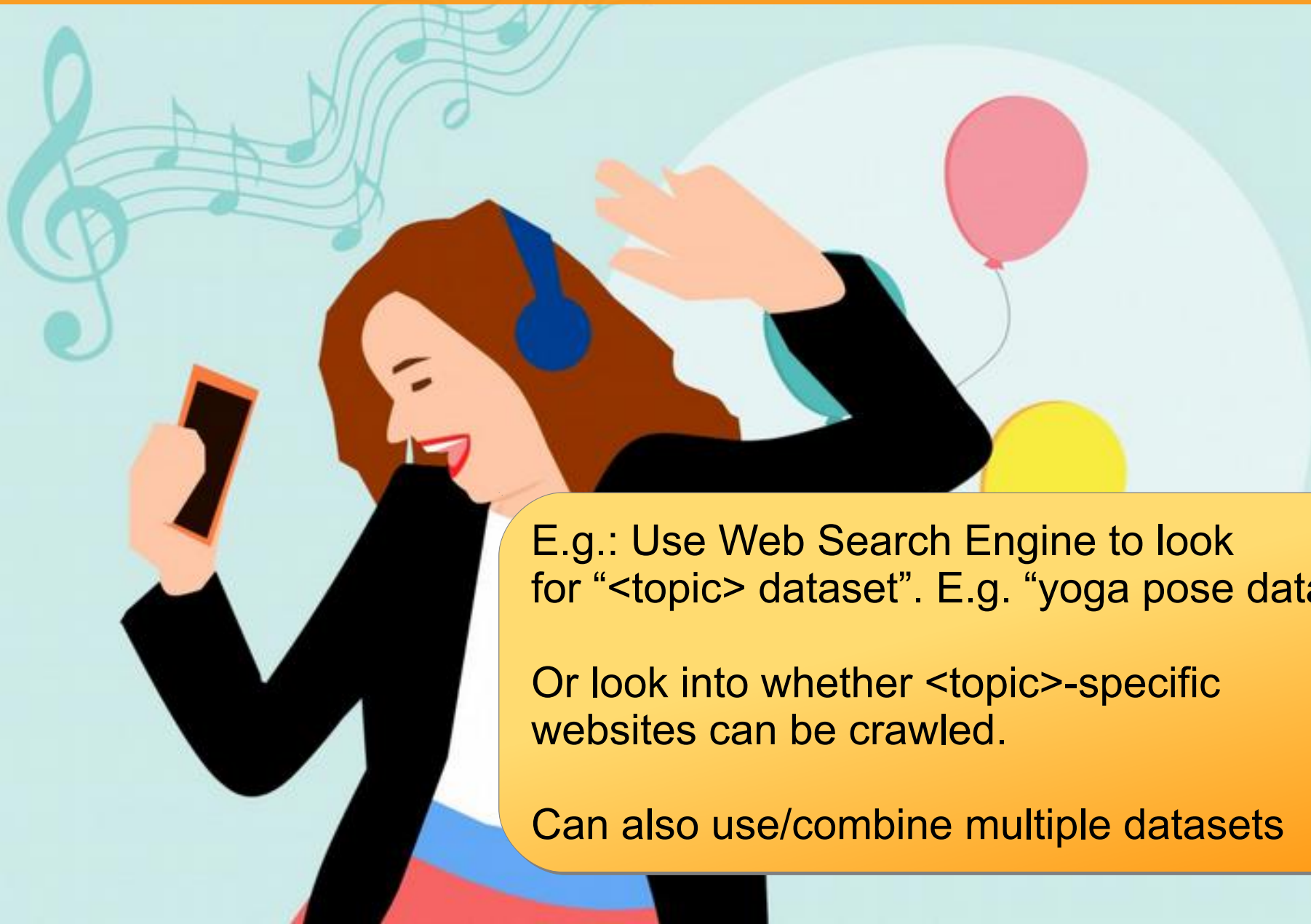
What are your hobbies?
What are you most
passionate about?

Music?
Travel?
Art?
Gaming?
etc.

Course Project



Course Project: Data



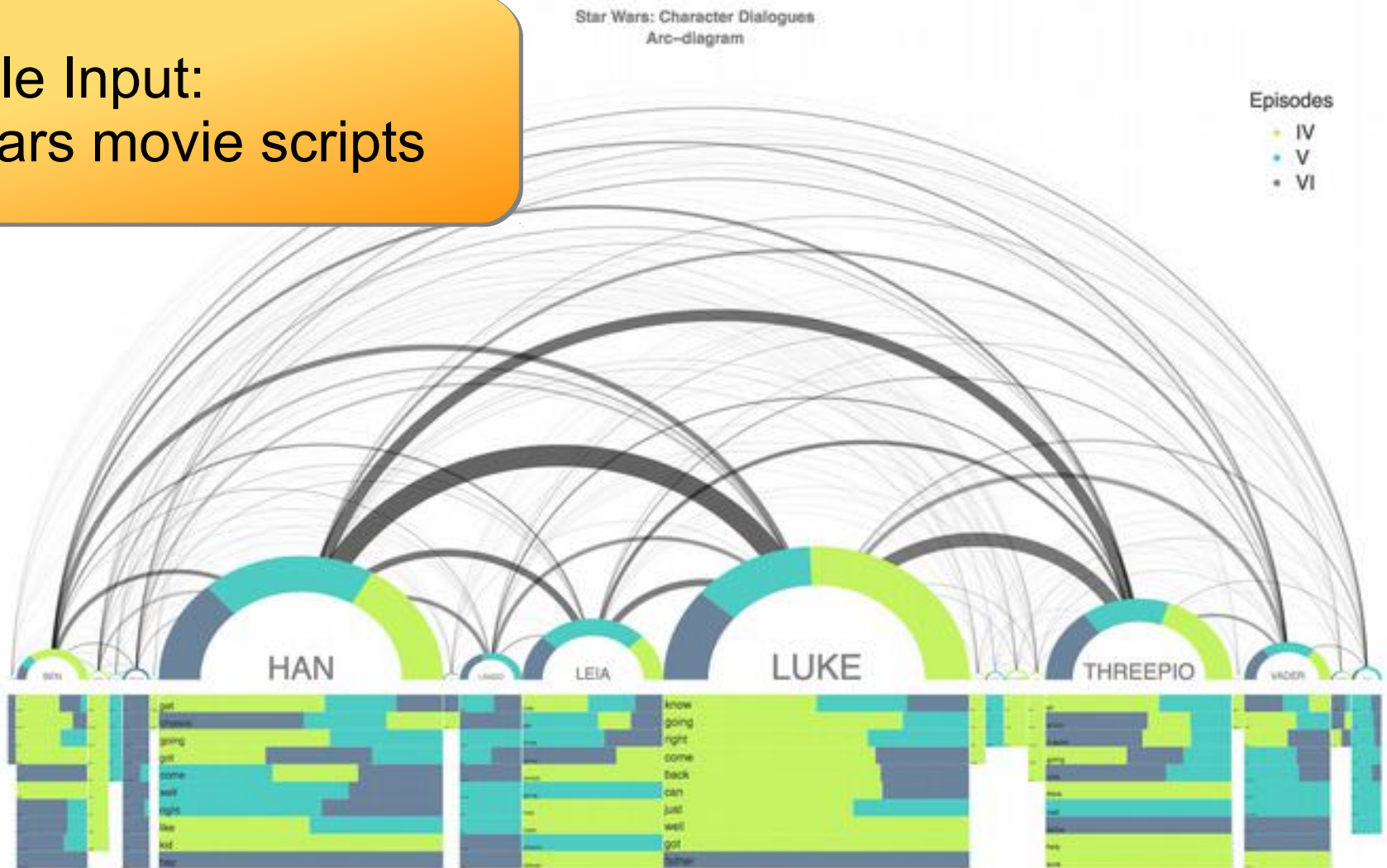
E.g.: Use Web Search Engine to look for “<topic> dataset”. E.g. “yoga pose dataset”

Or look into whether <topic>-specific websites can be crawled.

Can also use/combine multiple datasets

Course Project: Data

Example Input:
Star Wars movie scripts



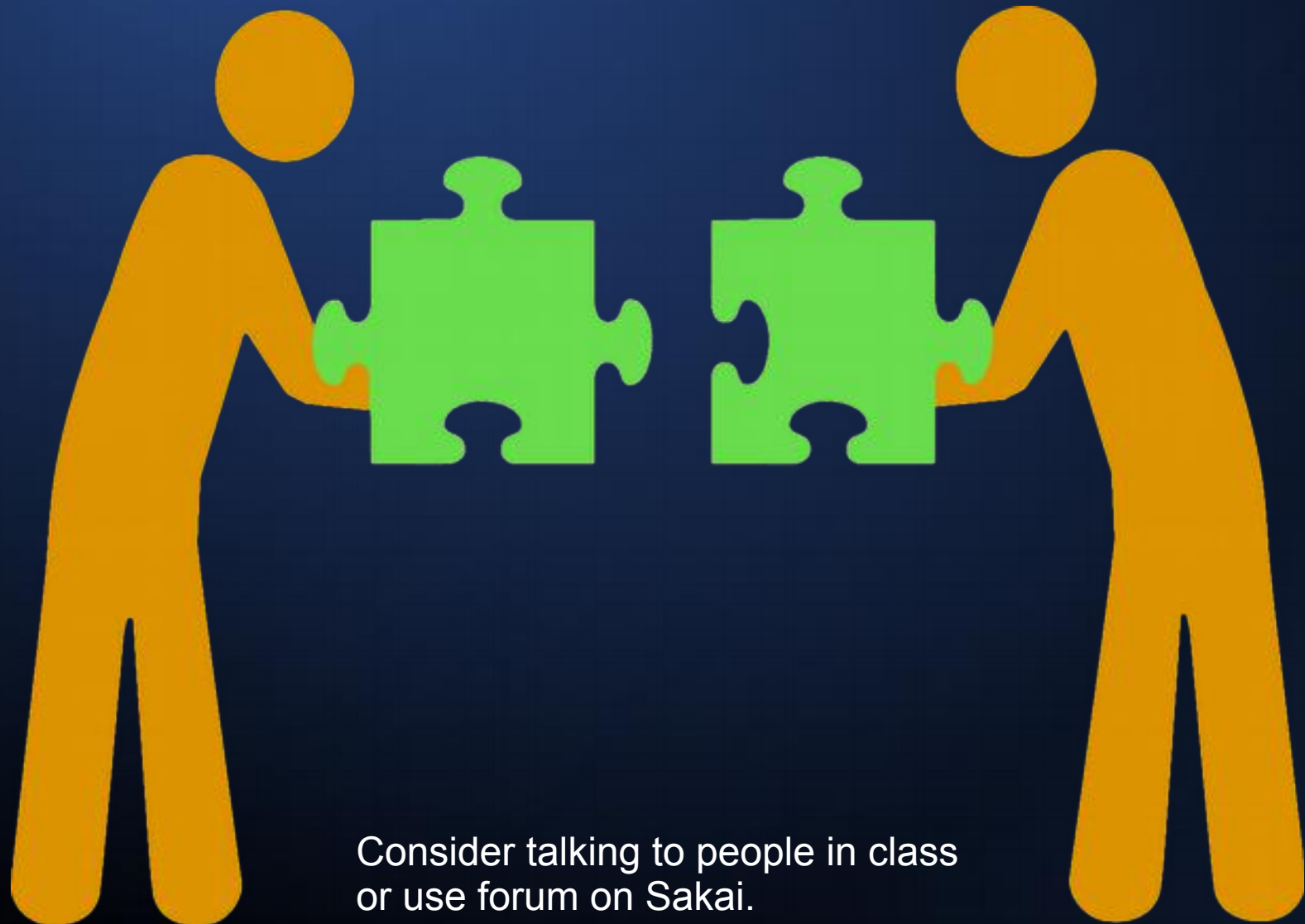
Course Project

Step 1: Short Project Proposal (by Feb. 17)

Just a 2-3 paragraph description of what you are planning to do and hoping to achieve, what data sources you hope to use, etc.

This can still be changed later, with approval from us.

Course Project: Team?



Consider talking to people in class
or use forum on Sakai.

Course Project: Team?

Teams

Team Size: 1 or 2

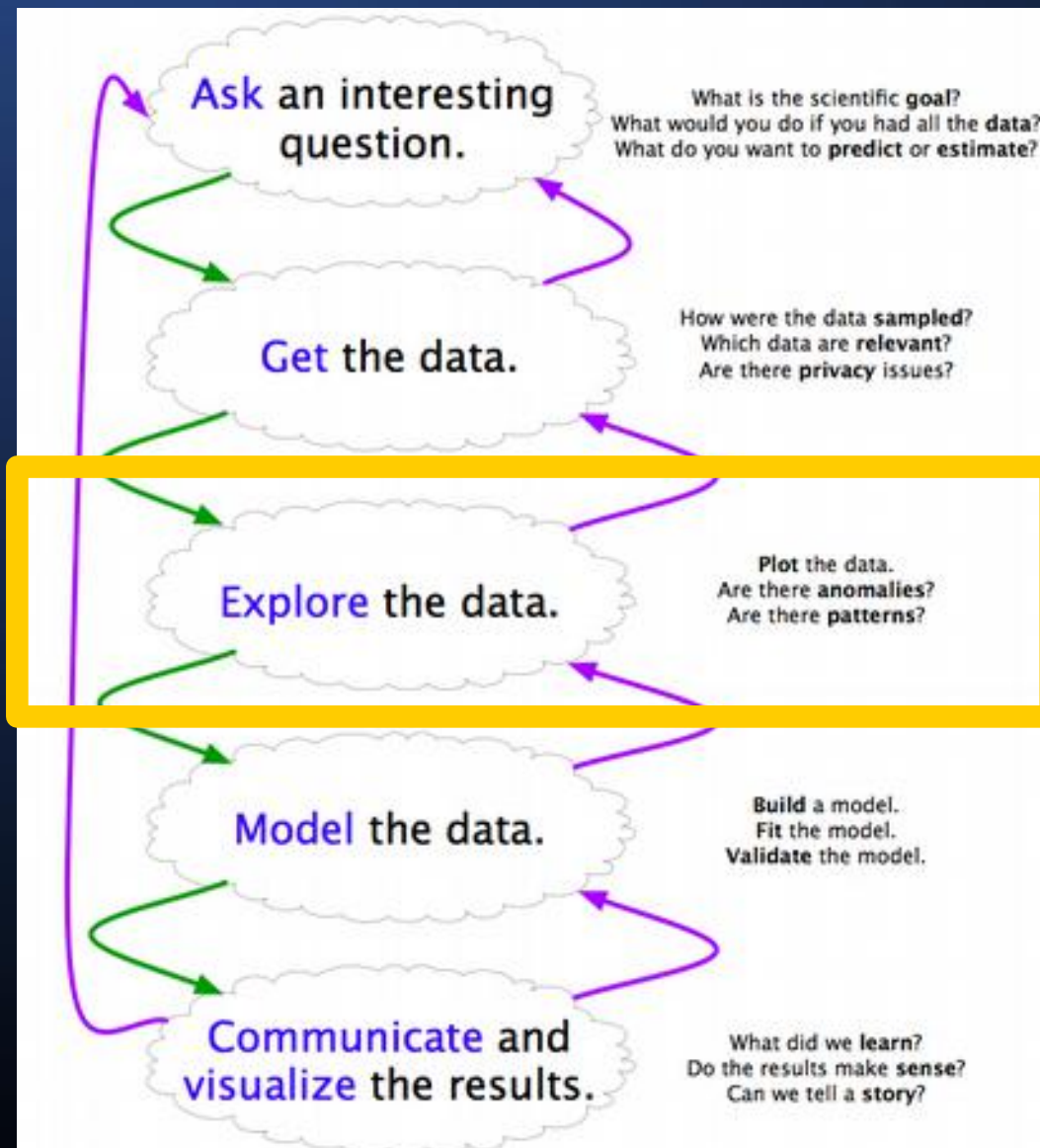
(exceptions only with prior approval,
for particularly large/challenging projects)

**Teams normally should not change after
submitting the proposal.**

Grade: Equal for all team members

However, we reserve the right to deviate from this
if the contributions were particularly unequal.

Course Project



Course Project: Deliverables

Intermediate Milestones as Homework

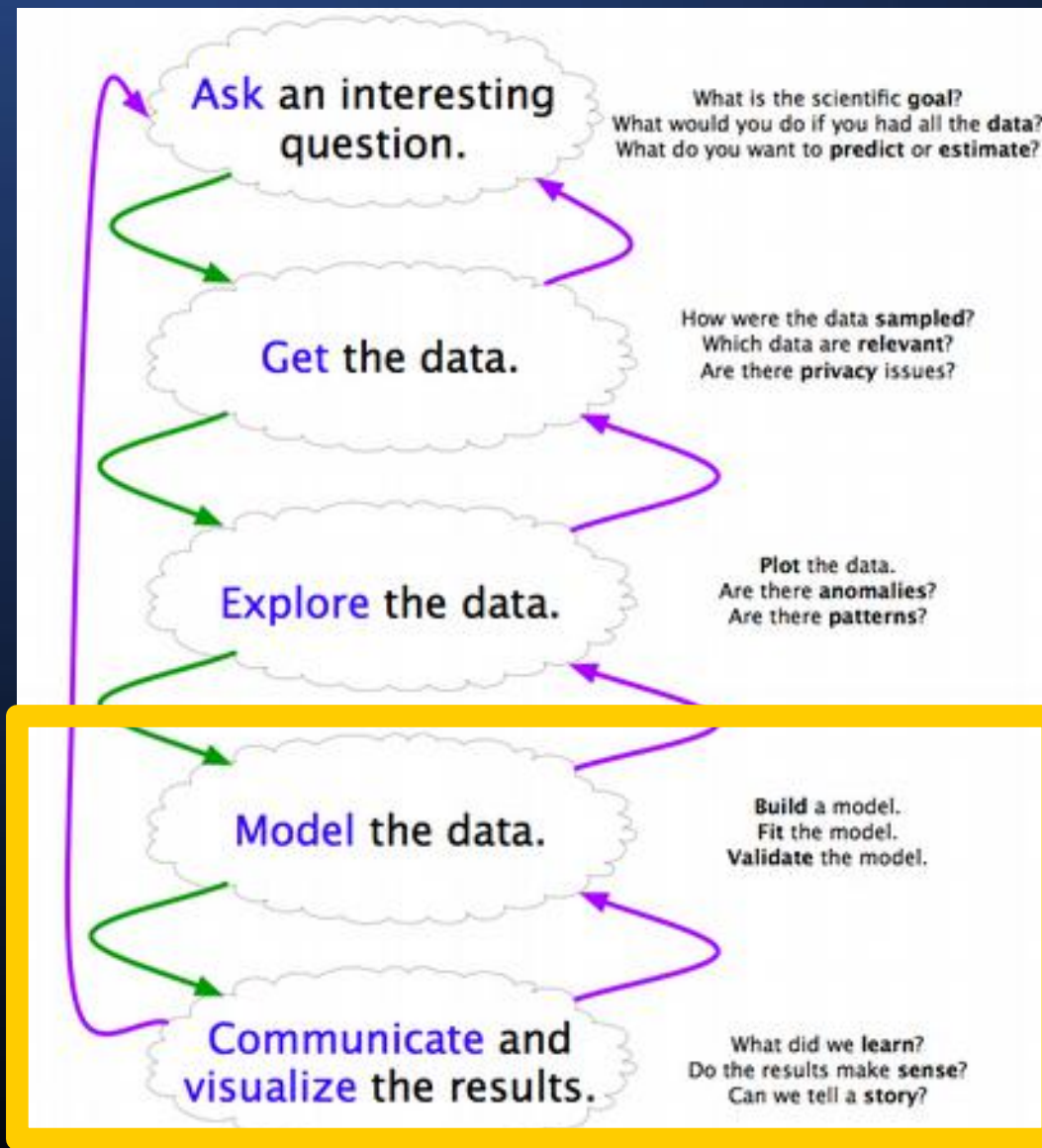
Dates: TBA

E.g.:

Describe data collection/source of data,
data format, data preprocessing.

Conduct data analysis, create visualizations

Course Project



Course Project: Deliverables

Short Project Presentations (Likely Apr. 28/30)

Very short (5-10 minute) presentation
of your work – probably not for all groups

May use slides or interactively demonstrate
your system.

Course Project: Deliverables

Final Report (by May 11)

1. Should include some machine learning experiments. Could also be a practical application.
2. Explain and evaluate your results, numerically or via visualizations. Should show what insights have been gained, or how well your method works.

Course Project: Deliverables

Final Report

3. Describe related work

Describe in what ways your work is different from what others have done online in blog posts/Kaggle/etc.

4. Conclusion section

What insights did you gain? What worked, what didn't work?

What else would you do if you had more time (or could start over)?

6. Acknowledgments section (mandatory!)

Mention all libraries used, all third-party material used!

You may not use third-party images without attribution.

Course Project: Academic Dishonesty

I will find you...

And I will report you!



Course Project: Details

Rules

You may use any external libraries, as long as you explicitly mention this in an “Acknowledgments” section in your report.

Any third-party material used, even if modified or translated from a different programming language, must be mentioned in the “Acknowledgments” section in your report. Clearly indicate the extent of your own contribution.

Course Project: Details

Submission Details

All deadlines refer to 11:59pm Eastern time.

Late submissions at discretion of instructor, but with grade penalty.

- 1) Include Source Code as Jupyter Notebook/ZIP file/etc.**
- 2) Please do not submit large original input data files.**

Course Project: What to Focus on

Key Elements for a Good Grade

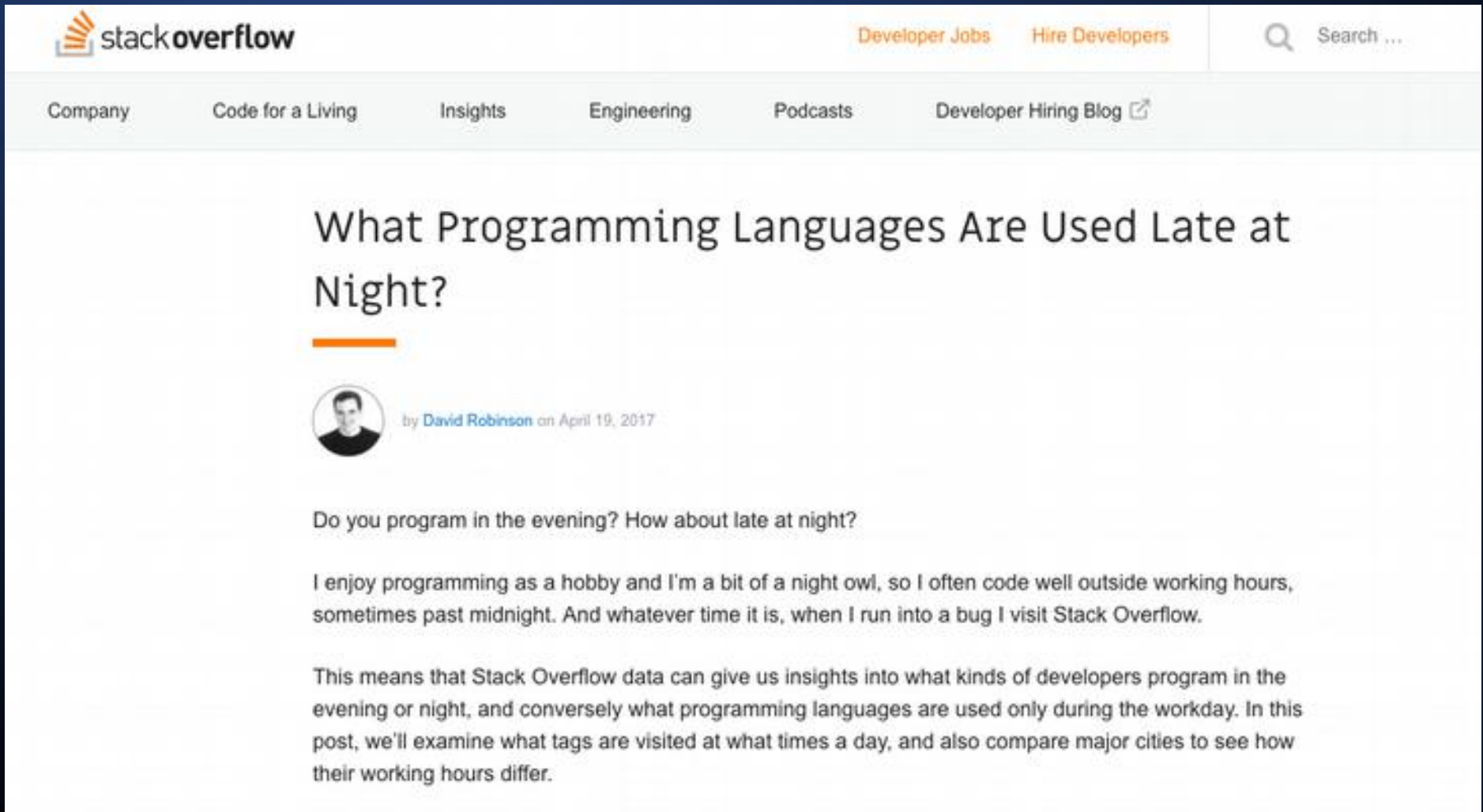
Put in effort

Sophisticated analysis methods: machine learning, etc. Beautiful visualizations are also nice, but primarily if they support analysis and insights about data.

Interesting idea:

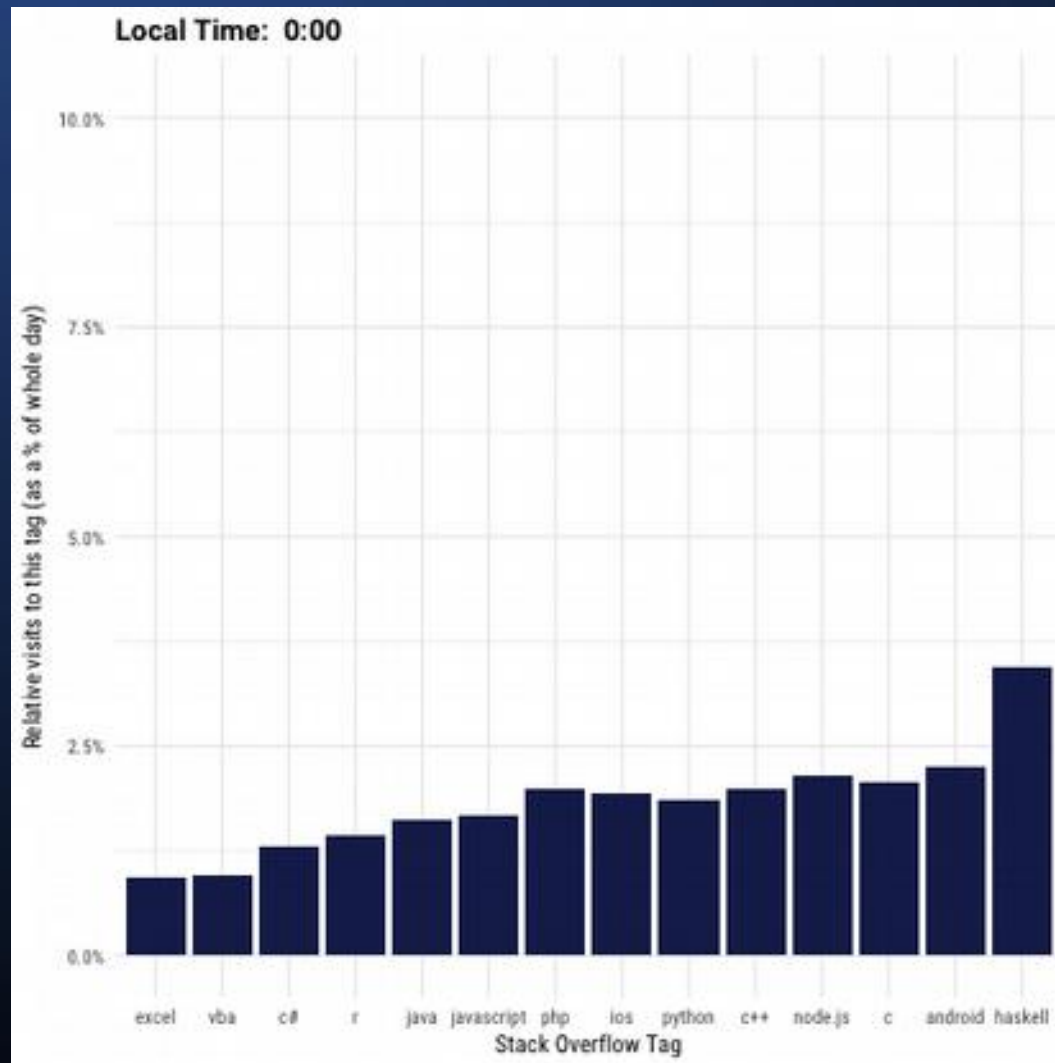
Imagine something that would score highly on Reddit or even that newspapers would write about

Course Project: Topic Examples (just for Inspiration)



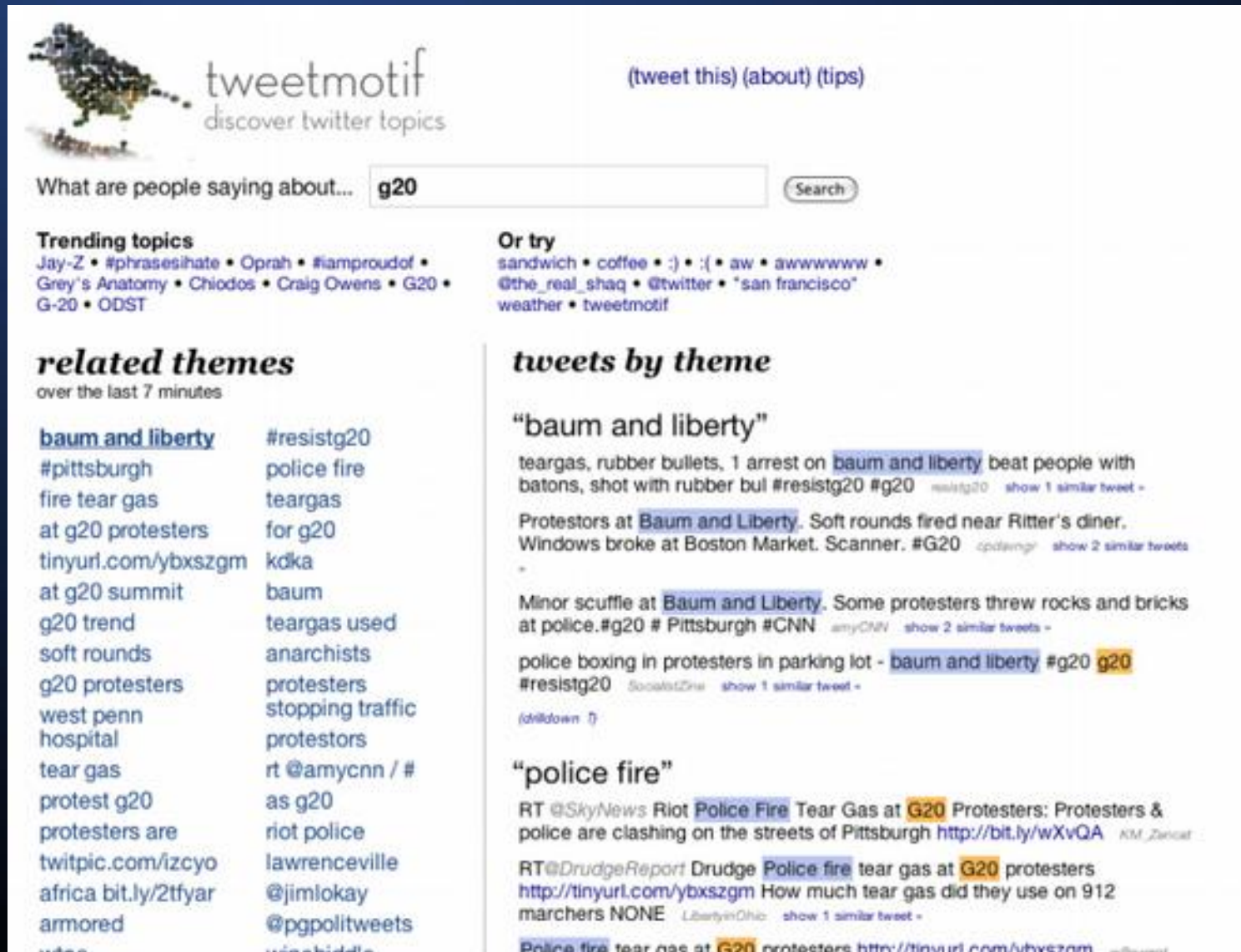
<https://stackoverflow.blog/2017/04/19/programming-languages-used-late-night/>

Course Project: Topic Examples (just for Inspiration)



<https://stackoverflow.blog/2017/04/19/programming-languages-used-late-night/>

Course Project: Topic Examples (just for Inspiration)



The screenshot shows the TweetMotif website, which is designed to help users discover Twitter topics. At the top, there is a logo for 'tweetmotif' with the tagline 'discover twitter topics'. Below the logo, there is a search bar with the text 'What are people saying about...' and the search term 'g20'. To the right of the search bar is a 'Search' button. Above the search bar, there are links for '(tweet this)', '(about)', and '(tips)'. Below the search bar, there are two sections: 'Trending topics' and 'Or try'. The 'Trending topics' section lists various topics including Jay-Z, #phrasesihate, Oprah, #iamproudof, Grey's Anatomy, Chiodos, Craig Owens, G20, G-20, and ODST. The 'Or try' section lists various topics including sandwich, coffee, :, {, aw, awwwwww, @the_real_shaq, @twitter, "san francisco", weather, and tweetmotif. Below these sections, there are two main columns of results. The left column is titled 'related themes' and lists various themes related to the search term 'g20', such as 'baum and liberty', '#resistg20', '#pittsburgh', 'police fire', 'fire tear gas', 'teargas', 'at g20 protesters', 'for g20', 'tinyurl.com/ybxszm', 'koka', 'at g20 summit', 'baum', 'g20 trend', 'teargas used', 'soft rounds', 'anarchists', 'g20 protesters', 'protesters', 'west penn', 'stopping traffic', 'hospital', 'protestors', 'tear gas', 'rt @amycnn / #', 'protest g20', 'as g20', 'protesters are', 'riot police', 'twitpic.com/izcyo', 'lawrenceville', 'africa bit.ly/2tfyar', '@jimlokay', 'armored', '@pgpolitweets', and 'winchiddle'. The right column is titled 'tweets by theme' and shows a list of tweets related to the search term 'g20'. The first tweet is titled '"baum and liberty"' and describes a protest in Pittsburgh where protesters used tear gas and rubber bullets. The second tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The third tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The fourth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The fifth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The sixth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The seventh tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The eighth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The ninth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks. The tenth tweet is titled '"police fire"' and describes a riot in Pittsburgh where police used tear gas and protesters used rocks and bricks.

tweetmotif
discover twitter topics

(tweet this) (about) (tips)

What are people saying about...

Trending topics
Jay-Z • #phrasesihate • Oprah • #iamproudof • Grey's Anatomy • Chiodos • Craig Owens • G20 • G-20 • ODST

Or try
sandwich • coffee • :) • { • aw • awwwwww • @the_real_shaq • @twitter • "san francisco" • weather • tweetmotif

related themes
over the last 7 minutes

baum and liberty #resistg20
#pittsburgh police fire
fire tear gas teargas
at g20 protesters for g20
tinyurl.com/ybxszm koka
at g20 summit baum
g20 trend teargas used
soft rounds anarchists
g20 protesters protesters
west penn stopping traffic
hospital protestors
tear gas rt @amycnn / #
protest g20 as g20
protesters are riot police
twitpic.com/izcyo lawrenceville
africa bit.ly/2tfyar @jimlokay
armored @pgpolitweets
winchiddle

tweets by theme

"baum and liberty"
teargas, rubber bullets, 1 arrest on **baum and liberty** beat people with batons, shot with rubber bul #resistg20 #g20 [show 1 similar tweet](#)
Protestors at **Baum and Liberty**. Soft rounds fired near Ritter's diner. Windows broke at Boston Market. Scanner. #G20 [show 2 similar tweets](#)
Minor scuffle at **Baum and Liberty**. Some protesters threw rocks and bricks at police. #g20 # Pittsburgh #CNN [show 2 similar tweets](#)
police boxing in protesters in parking lot - **baum and liberty** #g20 **g20** #resistg20 [show 1 similar tweet](#)
[\(collapse\)](#)

"police fire"
RT @SkyNews Riot **Police Fire** Tear Gas at **G20** Protesters: Protesters & police are clashing on the streets of Pittsburgh <http://bit.ly/wXvQA> [KM_Zencat](#)
RT@DrudgeReport Drudge **Police fire** tear gas at **G20** protesters <http://tinyurl.com/ybxszm> How much tear gas did they use on 912 marchers NONE [show 1 similar tweet](#)
Police fire tear gas at **G20** protesters <http://tinyurl.com/ybxszm> [show 1 similar tweet](#)

Course Project:

Topic Examples (just for Inspiration)

Sample of Word Anomalies

The Bible (King James Edition); Anonymous / Various

Frequent: *unto, lord, isreal, shall, god, mores, jesus, david, offering, tabernacle*

Infrequent: *girl, boy, school, success, condition, listen, princess*

Wonderful Wizard of Oz; Baum, Frank

Frequent: *woodman, scarecrow, witch, tin, emerald, monkeys, kansas, brains, winged*

Infrequent: *mother, money, soul, natural*

White Fang; London, Jack

Frequent: *musher, beaver, sled, dogs, cherokee, snarl*

Infrequent: *letter, person, window, green, sweet, loved, party, paper*

The Republic; Plato

Frequent: *guardians, unjust, true, injustice, state, gymnastic, rulers, democractical*

Infrequent: *miss, girl, boy, prince*

Alice's Adventures In Wonderland; Carroll (C.L. Dodgson), Lewis

Frequent: *gryphon, turtle, caterpillar, mock, dodo, mouse, rabbit, hedgehog*

Infrequent: *death, country, happy, fair, common*

Origin of the Species; Darwin, Charles

Frequent: *species, varieties, subaerial, selection, sterility, plants, modification, forms, variability*

Infrequent: *person, government, love, thinking, god, evil, fire*

Communist Manifesto; Marx, Karl/Engels, Friedrich

Frequent: *bourgeois, proletariat, communists, antagonisms, revolutionising, socialism, production, class, feudal, reactionary, exploitation, conditions, crises*

Infrequent: *said, love, why, heart, mother, poor, felt*

Paradise Lost; Milton, John

Frequent: *wonderous, heaven, satan, dominations*

Infrequent: *country, church, horses, sister*

Apology; Plato

Frequent: *corrupter, accusers, demigods, socrates, oracle, indictment*

Infrequent: *she, work, morning, replied, body*

Gargantua and Pantagruel; Rabelais, Francis

Frequent: *codpiece, catchpole, ballocks, dingdong, fart, chitterlings, gymnast, arse*

Infrequent: *smile, existence, feelings, british, professor, suffering*

Course Project: Topic Examples (just for Inspiration)

Mark Allen Thornton Home Research CV/Publications Blog **Software** Links

Welcome! This is a book recommender for the [Project Gutenberg](#) collection. If you type in the name of a book in the collection (**mostly pre-1930 texts**), it will suggest other books that have similar content or style.

If you already know one of the recommended books, you can rate how similar you think that book actually is to the one you were asking about. I'll use these ratings to help improve the recommendation system in the future.

Find books!

Showing best matches for [The Trial](#) by Franz Kafka:

| Content matches | | Style matches | |
|--|-------|--|-------|
| The Trial by Franz Kafka | ★★★★★ | The Trial by Franz Kafka | ★★★★★ |
| Burnham Breaker by Homer Greene | ★★★★★ | Shallow Soil by Knut Hamsun | ★★★★★ |
| The Quilt that Jack Built; How He Won the Bicycle by Annie Fellows Johnston | ★★★★★ | Sister Carrie by Theodore Dreiser | ★★★★★ |
| Flip's "Islands of Providence" by Annie Fellows Johnston | ★★★★★ | A Rogue by Compulsion by Victor Bridges | ★★★★★ |

Course Project: Topic Examples (just for Inspiration)

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Johnny Depp - Wikipedia



Wikipedia Article
→ **PPT**

Course Project: Topic Examples (just for Inspiration)

RUTGERS

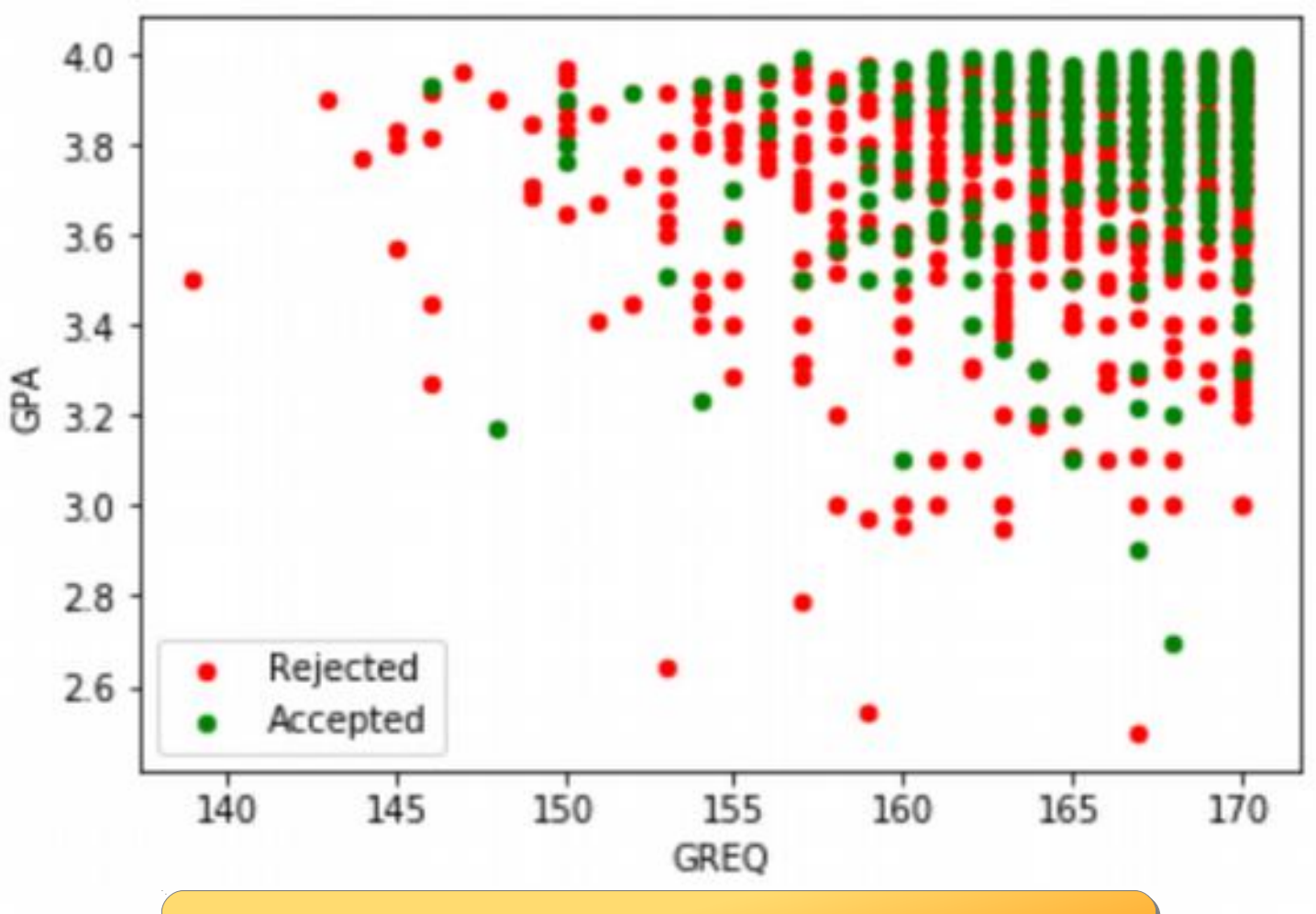
Early life

- Depp is of mostly English ancestry, with some ancestors from elsewhere in Europe.
- Depp is a 20th cousin of Elizabeth II.
- Depp moved frequently during his childhood.

**Related Topics
Available. Get in
touch with me.**

**Wikipedia Article
→ PPT**

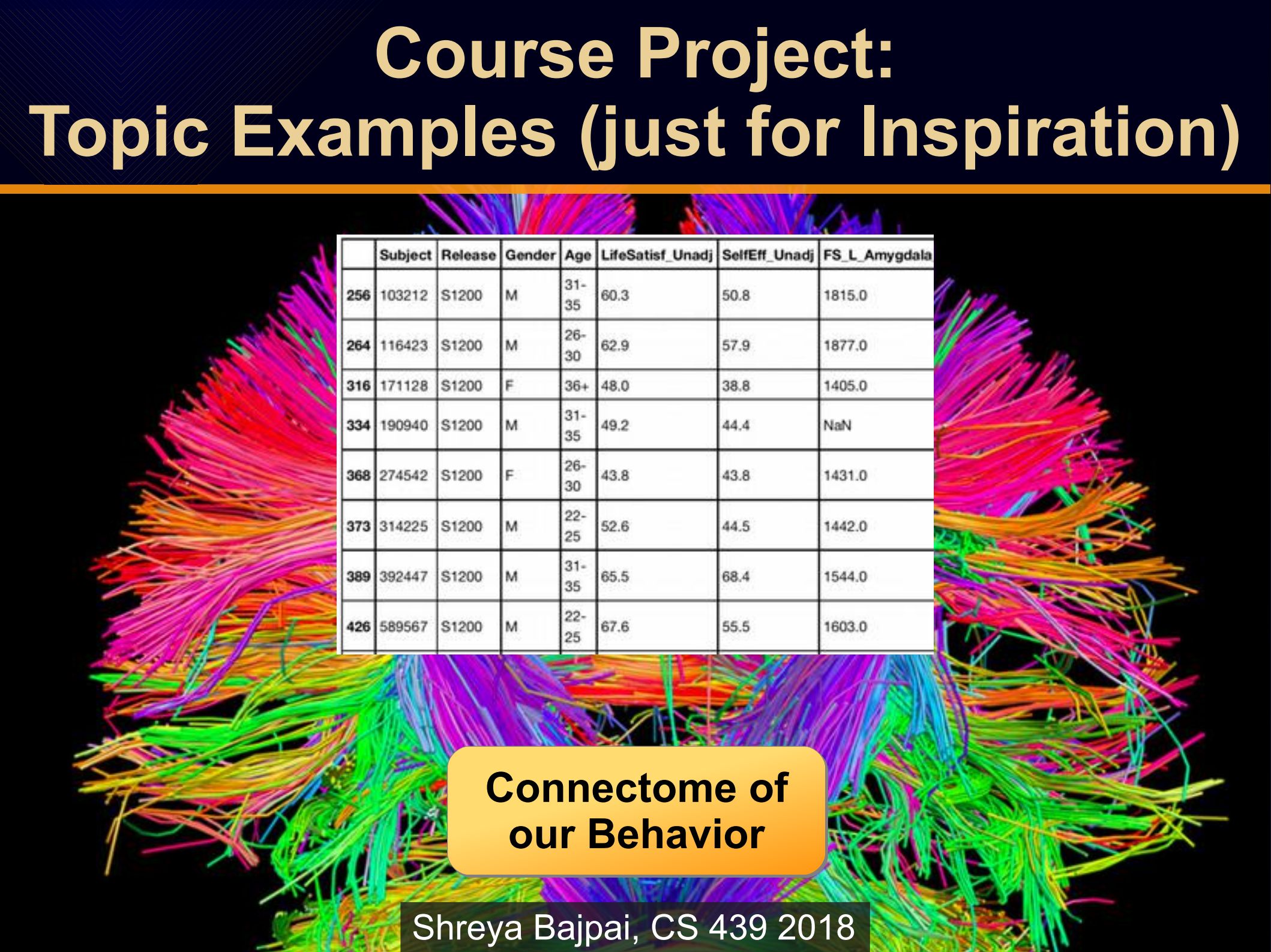
Topic Examples (just for Inspiration)



College Acceptance Statistics

Nithin Tammishetti, CS 439 2018

Course Project: Topic Examples (just for Inspiration)



| | Subject | Release | Gender | Age | LifeSatisf_Unadj | SelfEff_Unadj | FS_L_Amygdala |
|-----|---------|---------|--------|-------|------------------|---------------|---------------|
| 256 | 103212 | S1200 | M | 31-35 | 60.3 | 50.8 | 1815.0 |
| 264 | 116423 | S1200 | M | 26-30 | 62.9 | 57.9 | 1877.0 |
| 316 | 171128 | S1200 | F | 36+ | 48.0 | 38.8 | 1405.0 |
| 334 | 190940 | S1200 | M | 31-35 | 49.2 | 44.4 | NaN |
| 368 | 274542 | S1200 | F | 26-30 | 43.8 | 43.8 | 1431.0 |
| 373 | 314225 | S1200 | M | 22-25 | 52.6 | 44.5 | 1442.0 |
| 389 | 392447 | S1200 | M | 31-35 | 65.5 | 68.4 | 1544.0 |
| 426 | 589567 | S1200 | M | 22-25 | 67.6 | 55.5 | 1603.0 |

**Connectome of
our Behavior**

Shreya Bajpai, CS 439 2018

Course Project: Topic Examples (just for Inspiration)



Course Project



**Good course projects
can go on your CV!**