What Makes A Survey *☆*Useful*☆* For Demographics

Jessie Christiansen
Caltech/IPAC-NASA Exoplanet Science Institute

Overview

01
Why?

02
What?

03
Do's

04
DON'TS!

05
Example

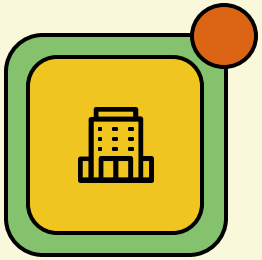
06
Resources

01

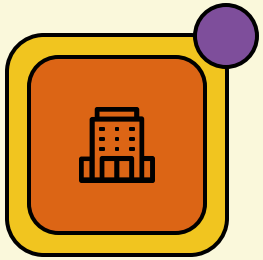
Why

...should you listen to this talk?

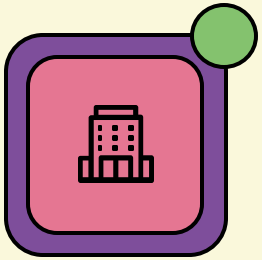
Do you...



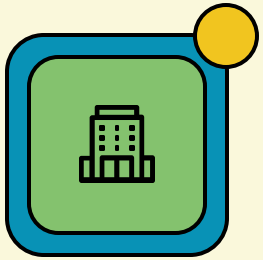
Want to perform a survey?



Want to assess if a survey is suitable for your needs?



Want to know how <u>not</u> to use data?



Referee papers on any of the above?

02

What

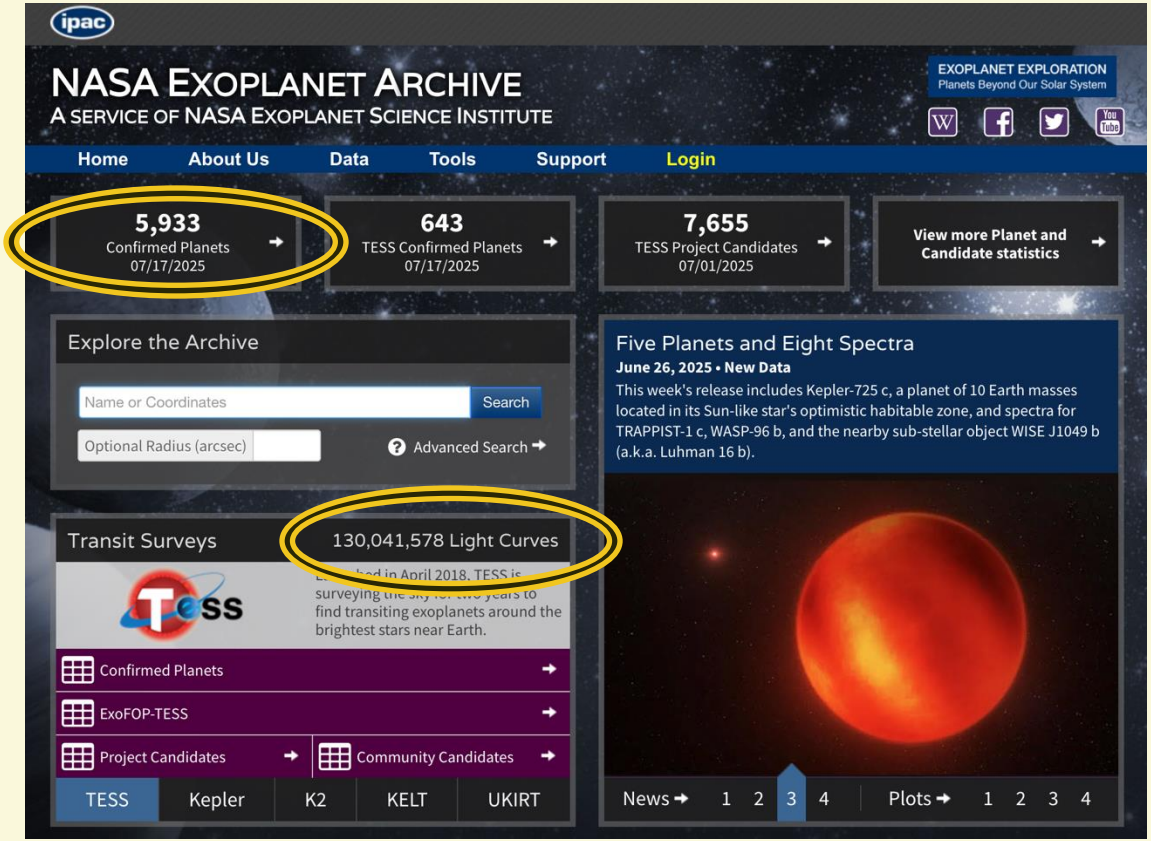
...is the problem we're trying to solve?

Astro2020, Goal 1:

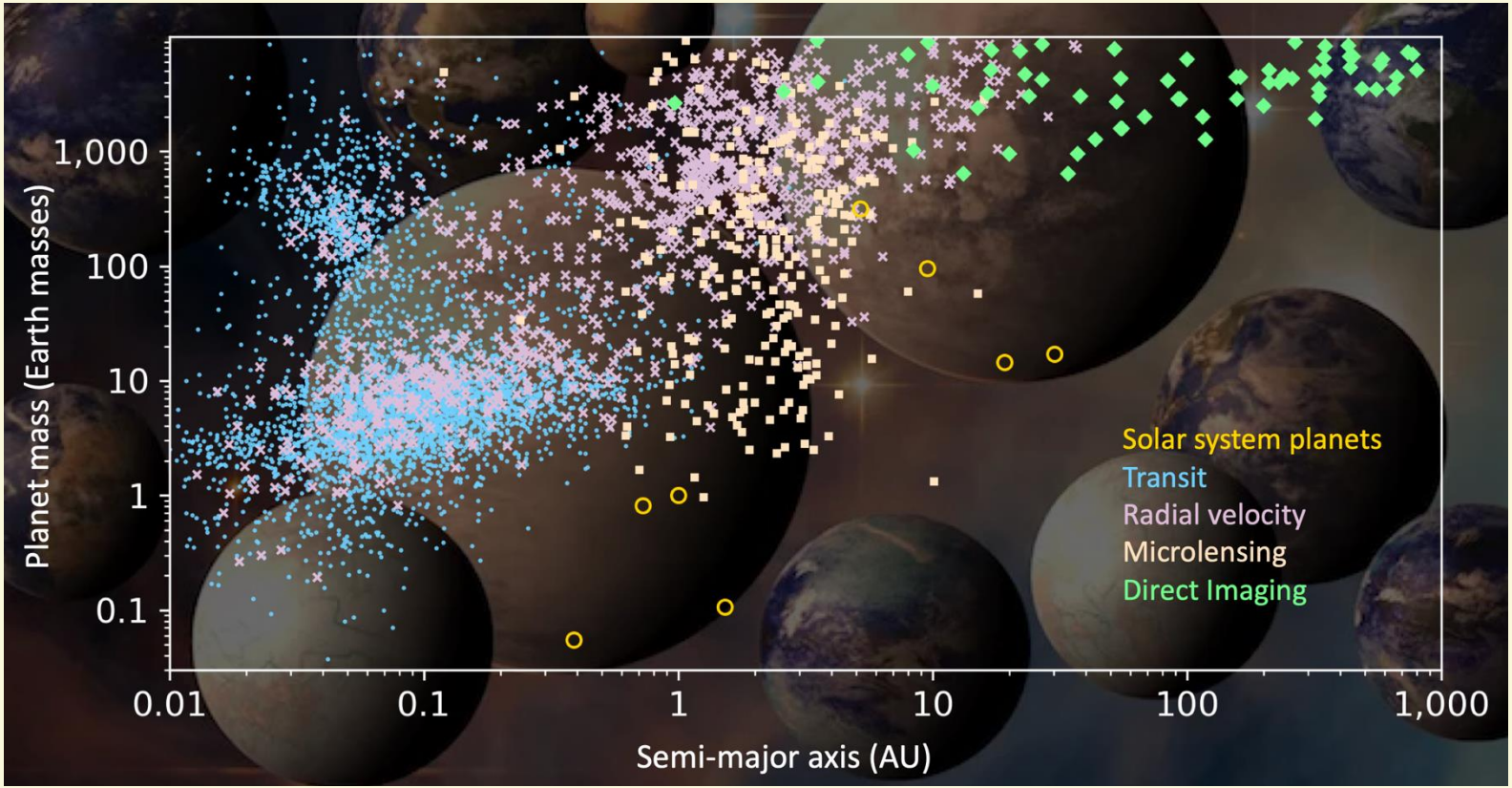
“to understand the formation and evolution of planetary systems as products of the process of star formation, and characterize and explain the diversity of planetary system architectures, planetary compositions, and planetary environments produced by these processes”

Finding: "*Current knowledge of the demographics and characteristics of planets and their systems is substantially incomplete.”*

The problem is not that we don't have enough data...



The problem is not that we don't have enough data...




Why don't we have fantastic demo-graphics yet?
Not enough meta-data!
Model/methodology differences/assumptions
Missing data, inc. provenance
Sample sizes
No selection criteria to cross-calibrate samples
No false positive/ negative rates
No detection criteria/ noise model provided
Calibration differences
GAME SHOW

03

Do's

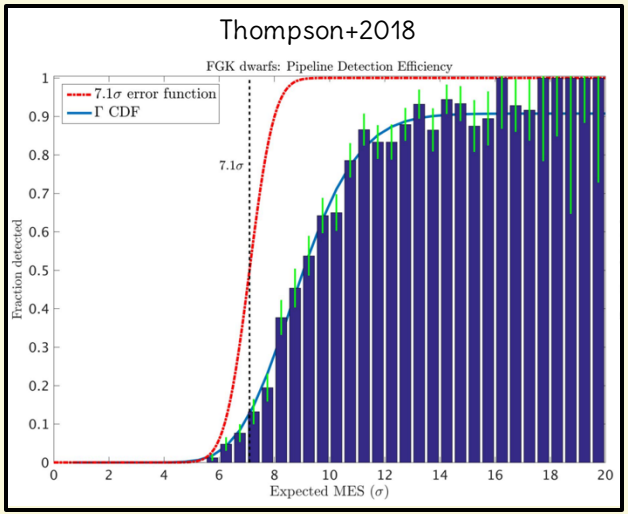
... in creating or evaluating a survey for demographics

(https://arxiv.org/pdf/2304.12442)

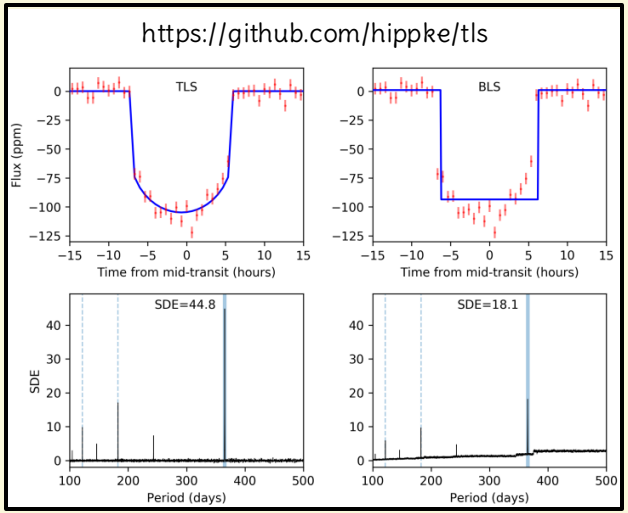
Does the survey have:

Survey properties	Stated detection criteria (e.g. SNR threshold, number of transits, etc?) Which models are being fit to the data? What is the observing baseline? Is it the same for every star or does it vary? A quantified method for handling binaries/multiple star systems? A DETECTION EFFICIENCY (false negative rate), and how is it measured?

Thompson+2018

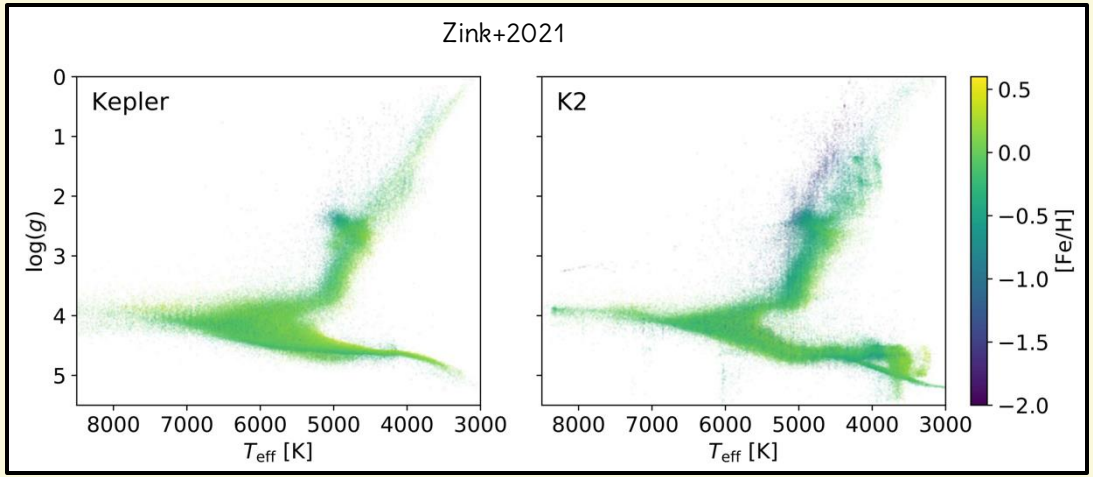


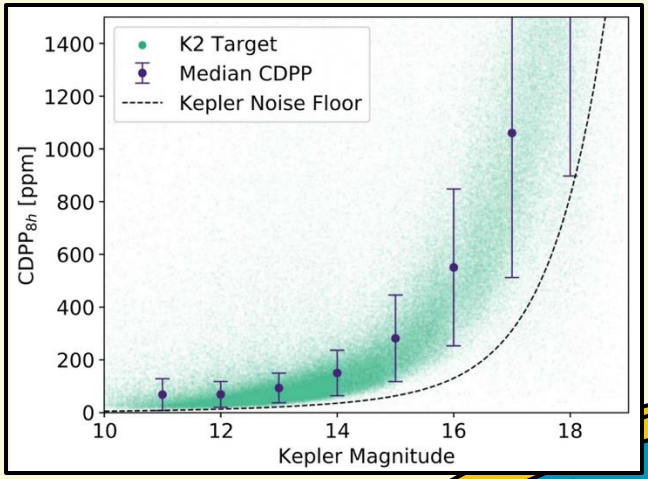
https://github.com/hippke/tls



Does the survey have:

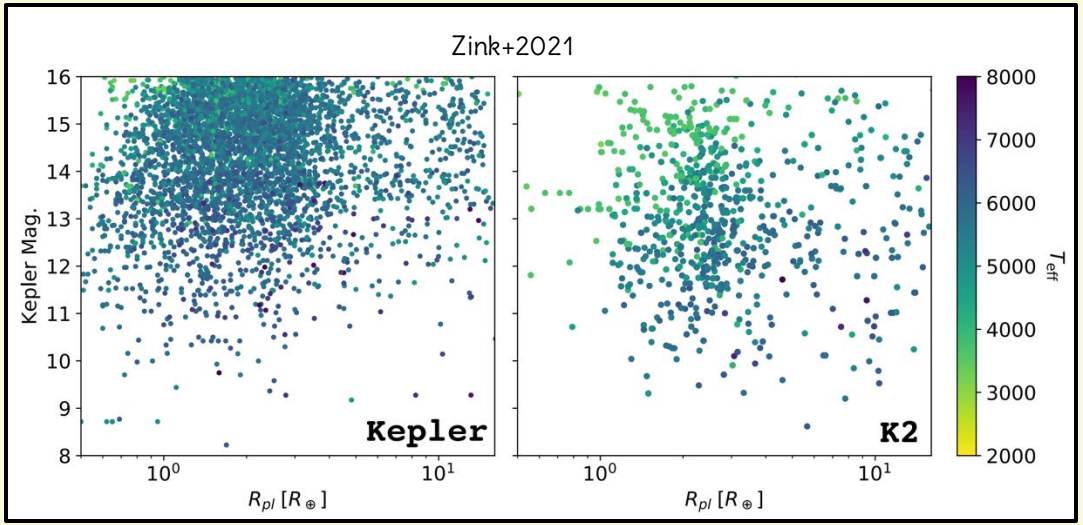
Stellar sample	A large enough sample of stars, and clear selection criteria for those stars? Which stellar parameters are used, and is their provenance UNIFORM? Which stars are removed from the final sample and why? Quantified noise properties for the stellar sample? IF this is the only data used, are there sufficient targets for your science?

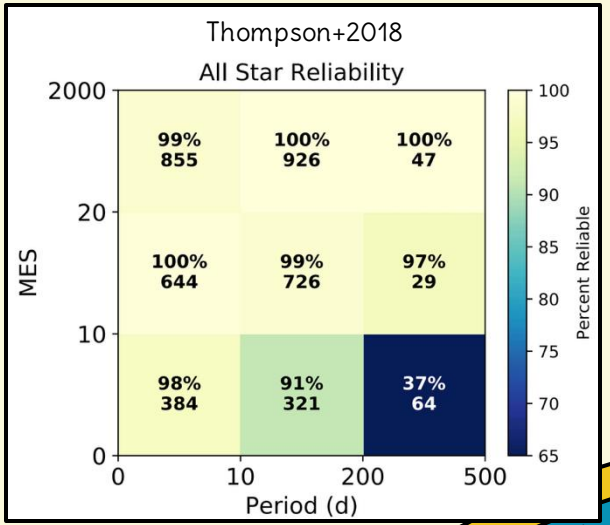




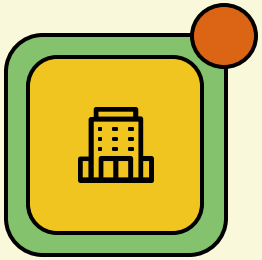
Does the survey have:

Planet catalog	Measured planet parameters, and how are they obtained? UNIFORMLY derived planet parameters, and what models are used? A catalog RELIABILITY? Is it measured or estimated? Does it include the astrophysical false positive and instrumental false alarm rates?



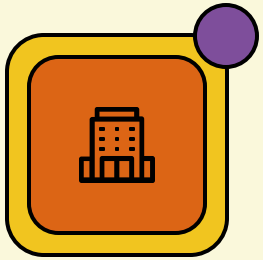


Are the data FAIR?



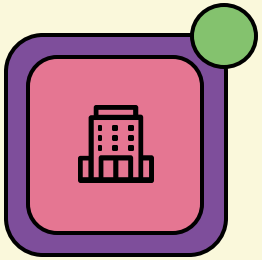
Findable

Metadata and data should be easy to find for both humans and computers.



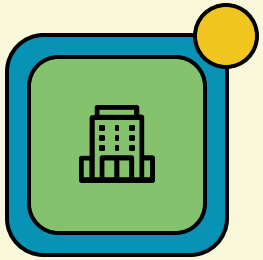
Accessible

(Meta)data are retrievable by their identifier using a standard communications protocol; metadata are available, even when the data no longer are



Interoperable

(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation



Reusable

Metadata and data should be well-described so that they can be replicated and/or combined in different settings, including a clear data usage license

https://www.go-fair.org/fair-principles/

04 DON'TS!

... in creating or evaluating a survey for demographics


NO
MANUAL
VETTING

DON'TS

... host the data on your institution's server. The server will not outlast your survey's usefulness!

... invent a new data format or standard for your data. An appropriate, open, free one already exists!

... assume you know all the use cases for your data. Provide (machine-readable!) meta-data to support a variety of use cases!

... throw data away without quantifying (and explaining) how/why!

... publish your detections and sit on your non-detections!

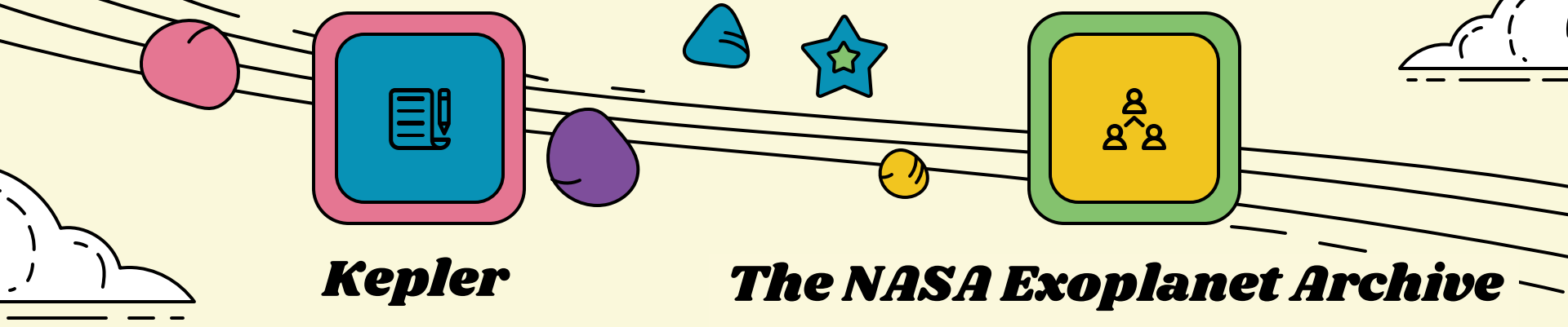
... your pet peeve here

05

Example

... in creating or evaluating a survey
for demographics

Which planet catalog should I use for demographics?



Kepler

The NASA Exoplanet Archive

Survey properties	✓	✗
Stellar sample	✓	✗
Planet catalog	✓	✗

Example survey design

Searching for Habitable Exoplanets with Relative Astrometry (SHERA)	How many widely separated (20-200au) binary stars do we need to survey for small, habitable zone planets to determine whether the occurrence rate is similar to single stars? (Important input to planet formation physics, and how to prioritize binary stars with future missions like Habitable Worlds Observatory)

t	***8***	***14***
r	***0.46***	***0.46***
k	***0***	***0***
P	***0.025***	***0.001***

r = occurrence rate

t = number of targets

k = number of planets detected

$$P(k \text{ events in interval } t) = \frac{(rt)^k e^{-rt}}{k!}$$

Formulate your hypothesis, examine whether the survey can test it

06

Resources

... for creating or evaluating a survey
for demographics

Resources

SIG#2 Report: Enabling Exoplanet Demographics Studies with Standardized Exoplanet Survey Meta-Data

NASA Exoplanet Archive Kepler Mission Page

https://www.go-fair.org/fair-principles/

Teske, Wang, Wolfgang et al. 2021; Batalha, Wolfgang, Teske et al. 2023

Conclusions

No manual vetting!

Utilize metadata!