# Playing in the Cosmic Backyard: A Statistician's Journey into Astronomy

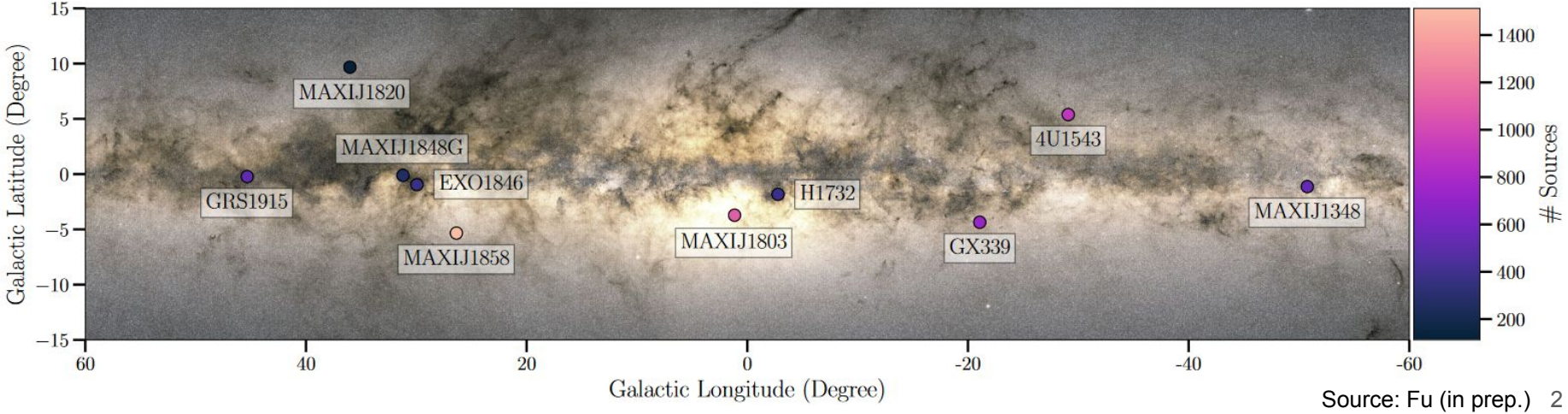
Shih Ching Fu
shihching.fu@postgrad.curtin.edu.au

ICRAR
International Centre for Radio Astronomy Research

Curtin University

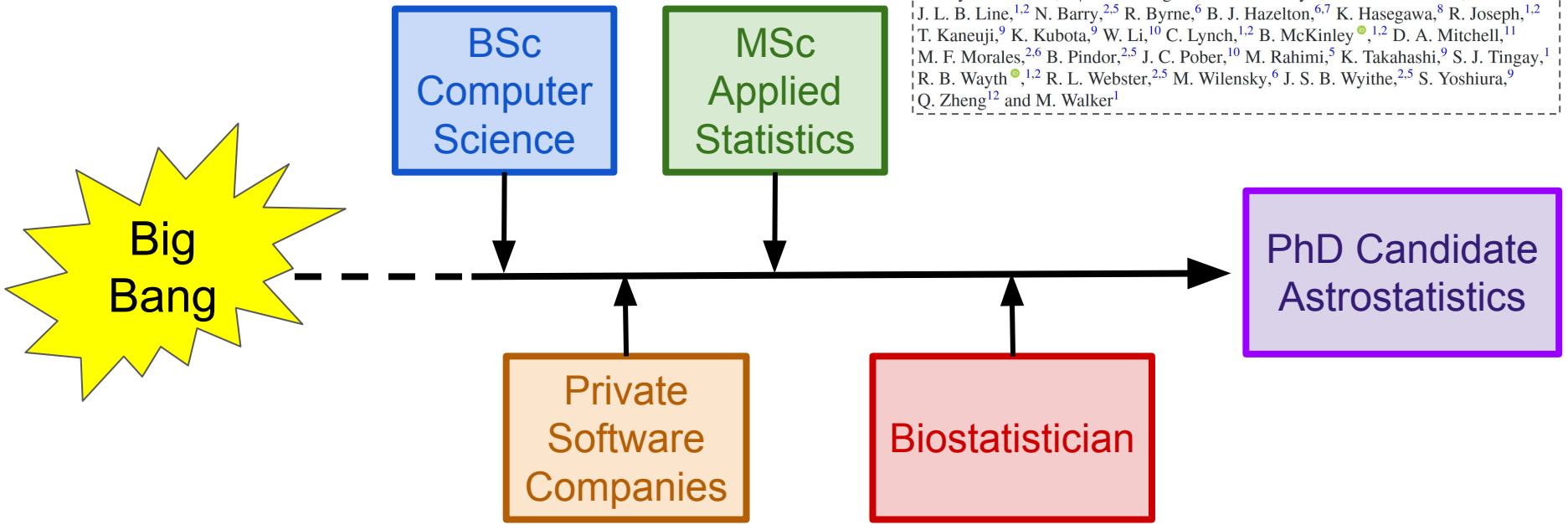
The best thing about being a statistician is that you get to play in everyone's backyard.

John Tukey




Source: Fu (in prep.)

# My journey so far…




Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS 486, 5766–5784 (2019) doi:10.1093/mnras/stz1207
Advance Access publication 2019 May 8

**Robust statistics towards detection of the 21 cm signal from the Epoch of Reionization**

Cathryn M. Trott,[1,2]† Shih Ching Fu,[3] S. G. Murray,[1,2,4] C. H. Jordan,[1,2] J. L. B. Line,[1,2] N. Barry,[2,5] R. Byrne,[6] B. J. Hazelton,[6,7] K. Hasegawa,[8] R. Joseph,[1,2] T. Kaneuji,[9] K. Kubota,[9] W. Li,[10] C. Lynch,[1,2] B. McKinley,[1,2] D. A. Mitchell,[11] M. F. Morales,[2,6] B. Pindor,[2,5] J. C. Pober,[10] M. Rahimi,[5] K. Takahashi,[9] S. J. Tingay,[1] R. B. Wayth,[1,2] R. L. Webster,[2,5] M. Wilensky,[6] J. S. B. Wyithe,[2,5] S. Yoshiura,[9] Q. Zheng[12] and M. Walker[1]

# A 12 parsec view of astrostatistics

1. Academic astronomers
2. Astronomical data
3. Astronomical statistics

Source: flickr

# Astrostatistics…

…and astroinformatics are **interdisciplinary** fields that perform research at the **interface of astronomy and statistics**, computer science, applied math, and data analytics…

Eadie et al. (2019)

… is the analysis of astronomical observations and **linking data to astrophysical theory**….

Feigelson et al. (2021)

# Academic Astronomers

- Quite numerate but statistics is not part of their typical training.
- Competent in general purpose languages, e.g., Python, C.
- *Very* large collaborations
  - International
  - Publicly funded
- Everything is on the arXiv.
- Diverse and welcoming community.

***Euclid*: Early Release Observations – The intracluster light and intracluster globular clusters of the Perseus cluster**★

M. Kluge[1], N. A. Hatch[★★2], M. Montes[3,4], J. B. Golden-Marx[2], A. H. Gonzalez[5], J.-C. Cuillandre[6], M. Bolzonella[7], A. Lançon[8], R. Laureijs[9], T. Saifollahi[8,10], M. Schirmer[11], C. Stone[12], A. Boselli[13,14], M. Cantiello[15], J. G. Sorce[16,17,18], F. R. Marleau[19], P.-A. Duc[20], E. Sola[21], M. Urbano[8], S. L. Ahad[22,23], Y. M. Bahé[24], S. P. Bamford[2], C. Bellhouse[2], F. Buitrago[25,26], P. Dimauro[27,28], F. Durret[29], A. Ellien[30], Y. Jimenez-Teja[31,28], E. Slezak[32], N. Aghanim[17], B. Altieri[33], S. Andreon[34], N. Auricchio[7], M. Baldi[35,7,36], A. Balestra[37], S. Bardelli[7], R. Bender[1,38], D. Bonino[39], E. Branchini[40,41,34], M. Brescia[42,43,44], J. Brinchmann[45,46], S. Camera[47,48,39], G. P. Candini[49], V. Capobianco[39], C. Carbone[50], J. Carretero[51,52], S. Casas[53], M. Castellano[27], S. Cavuoti[43,44], A. Cimatti[54], G. Congedo[55], C. J. Conselice[56], L. Conversi[57,33], Y. Copin[58], F. Courbin[24], H. M. Courtois[59], M. Cropper[49], A. Da Silva[60,61], H. Degaudenzi[62], J. Dinis[60,61], C. A. J. Duncan[56,63], X. Dupac[33], S. Dusini[64], M. Farina[65], S. Farrens[6], S. Ferriol[58], P. Fosalba[66,67], M. Frailis[68], E. Franceschi[7], M. Fumana[50], S. Galeotta[68], B. Garilli[50], W. Gillard[69], B. Gillis[55], C. Giocoli[7,70], P. Gómez-Alvarez[71,33], B. R. Granett[34], A. Grazian[37], F. Grupp[1,38], L. Guzzo[72,34], S. V. H. Haugan[73], J. Hoar[33], H. Hoekstra[74], W. Holmes[75], I. Hook[76], F. Hormuth[77], A. Hornstrup[78,79], P. Hudelot[80], K. Jahnke[11], E. Keihänen[81], S. Kermiche[69], A. Kiessling[75], T. Kitching[49], R. Kohley[33], B. Kubik[58], M. Kümmel[38], M. Kunz[82], H. Kurki-Suonio[83,84], O. Lahav[85], S. Ligori[39], P. B. Lilje[73], V. Lindholm[83,84], I. Lloro[86], E. Maiorano[7], O. Mansutti[68], O. Marggraf[87], K. Markovic[75], N. Martinet[13], F. Marulli[88,7,36], R. Massey[89], S. Maurogordato[32], H. J. McCracken[80], E. Medinaceli[7], S. Mei[90], M. Melchior[91], Y. Mellier[29,80], M. Meneghetti[7,36], E. Merlin[27], G. Meylan[24], M. Moresco[88,7], L. Moscardini[88,7,36], E. Munari[68,92], R. C. Nichol[93], S.-M. Niemi[9], J. W. Nightingale[94,95], C. Padilla[96], S. Paltani[62], F. Pasian[68], K. Pedersen[97], W. J. Percival[22,23,98], V. Pettorino[9], S. Pires[6], G. Polenta[99], M. Poncet[100], L. A. Popa[101], L. Pozzetti[7], G. D. Racca[9], F. Raison[1], R. Rebolo[3,4], A. Renzi[102,64], J. Rhodes[75], G. Riccio[43], H.-W. Rix[11], E. Romelli[68], M. Roncarelli[7], E. Rossetti[35], R. Saglia[38,1], D. Sapone[103], B. Sartoris[38,68], M. Sauvage[6], R. Scaramella[27,104], P. Schneider[87], T. Schrabback[19], A. Secroun[69], G. Seidel[11], M. Seiffert[75], S. Serrano[66,105,106], C. Sirignano[102,64], G. Sirri[36], J. Skottfelt[107], L. Stanco[64], P. Tallada-Crespí[51,52], A. N. Taylor[55], H. I. Teplitz[108], I. Tereno[60,26], R. Toledo-Moreo[109], F. Torradeflot[52,51], I. Tutusaus[110], E. A. Valentijn[10], L. Valenziano[7,111], T. Vassallo[38,68], G. Verdoes Kleijn[10], A. Veropalumbo[34,41,112], Y. Wang[108], J. Weller[38,1], O. R. Williams[113], G. Zamorani[7], E. Zucca[7], A. Biviano[68,92], C. Burigana[114,111], G. De Lucia[68], K. George[38], V. Scottez[29,115], P. Simon[87], A. Mora[116], J. Martín-Fleitas[116], F. Ruppin[58], and D. Scott[117]

*(Affiliations can be found after the references)*

November 18, 2024

# Are just as bad at naming things as statisticians

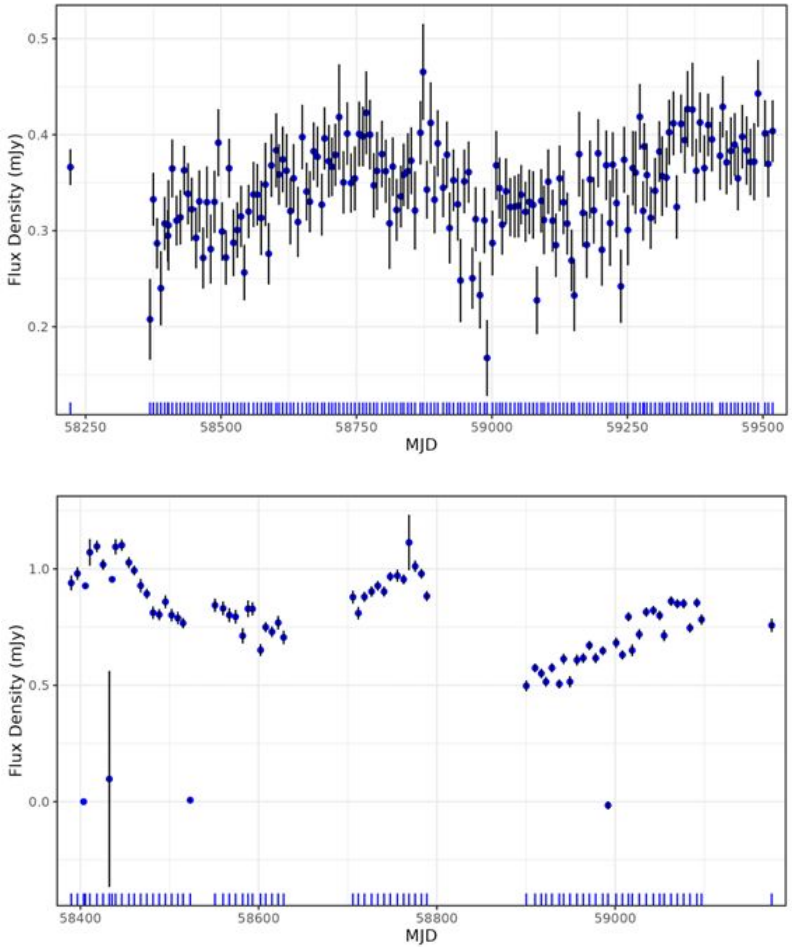
**Supernova taxonomy**[64][65]

<table>
<tr><td rowspan="5">Type I<br>No hydrogen</td><td colspan="3">Type Ia<br>Presents a singly ionised silicon (Si II) line at 615.0 nm (nanometers), near peak light</td><td>Thermal runaway</td></tr>
<tr><td rowspan="2">Type Ib/c<br>Weak or no silicon absorption feature</td><td colspan="2">Type Ib<br>Shows a non-ionised helium (He I) line at 587.6 nm</td><td rowspan="5">Core collapse</td></tr>
<tr><td colspan="2">Type Ic<br>Weak or no helium</td></tr>
<tr><td rowspan="3">Type II<br>Shows hydrogen</td><td rowspan="2">Type II-P/-L/n<br>Type II spectrum throughout</td><td rowspan="2">Type II-P/L<br>No narrow lines</td><td>Type II-P<br>Reaches a "plateau" in its light curve</td></tr>
<tr><td>Type II-L<br>Displays a "linear" decrease in its light curve (linear in magnitude versus time)[66]</td></tr>
<tr><td colspan="3">Type IIn<br>Some narrow lines</td></tr>
<tr><td colspan="4">Type IIb<br>Spectrum changes to become like Type Ib</td></tr>
</table>

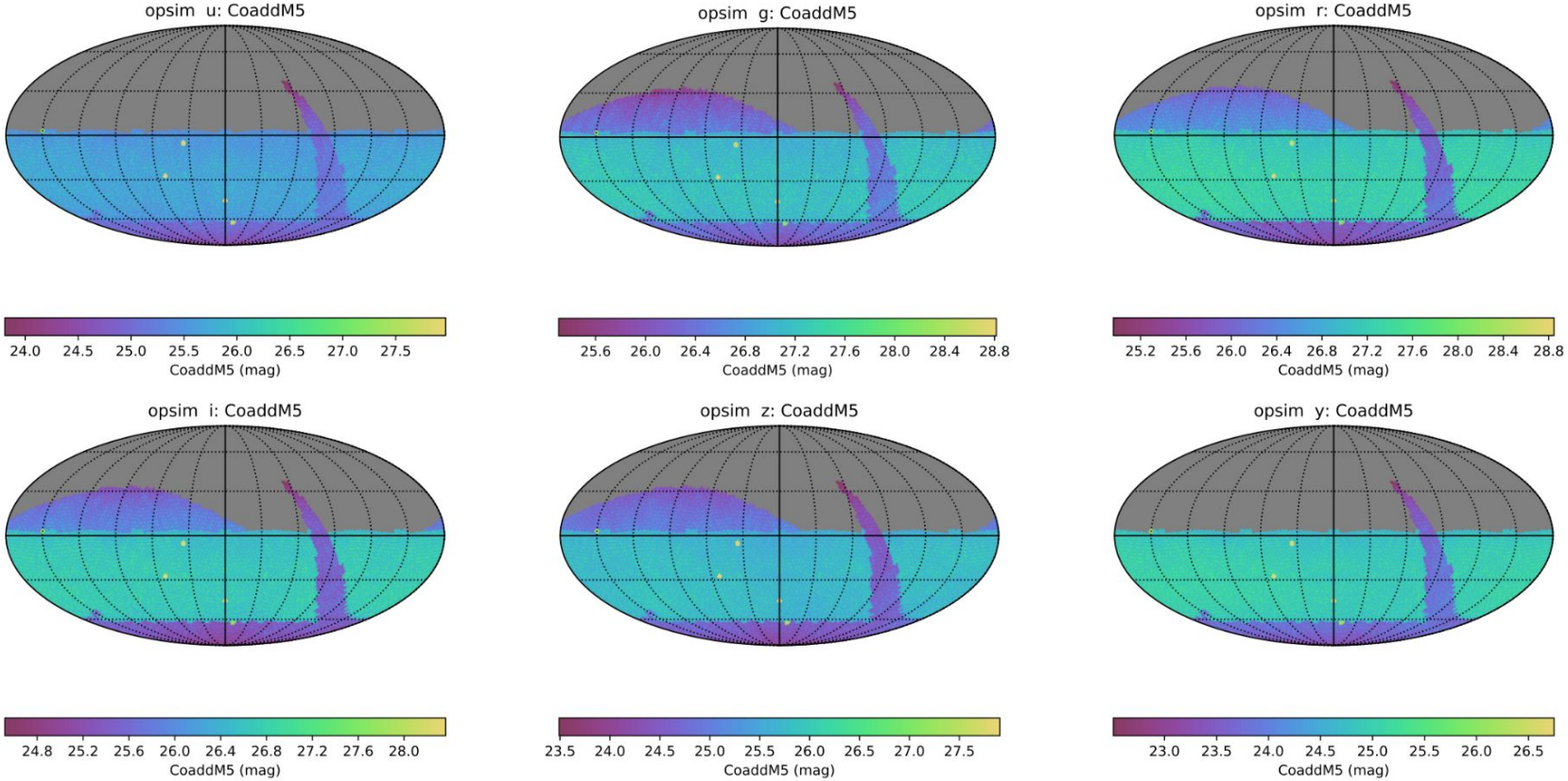
Source: Wikipedia

# Astronomical Data

- Observational data
- Dozens to millions of points
- Quality is varied
- Data includes error estimates
  - heteroscedastic
- Selection bias
- Missingness is NAR
- Non-gaussianity everywhere



# opsim: baseline2018a



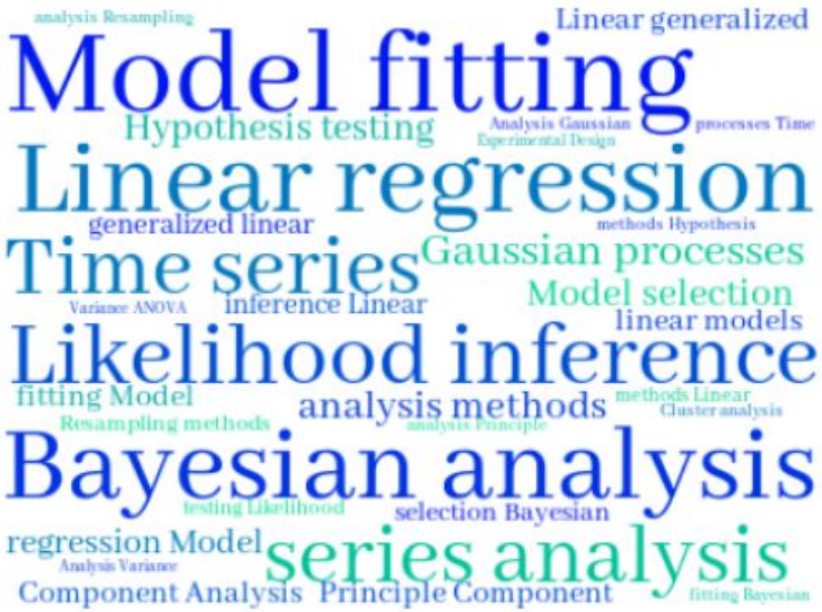

Source: [The LSST Operations Simulation Team (2018)]

# Statistical Practice in Astronomical Science

- Hold very tightly to mechanistic models and physically inspired parameterisations.
- Interested in explanatory models rather than predictive models.
- Very active practitioners of Bayesian inference.
- Often confine themselves to narrow suite of familiar methods.
- Like to try cutting edge tools, e.g., deep learning.

> Never tell me the odds.
>
> Han Solo

What are the statistical tools/models you find useful in your work?




What statistical tools/models do you wish you knew more about?


Source: Eadie et al. (2019)

# The Astrostatistics Community

- International Astrostatistics Association (IAA)
- Astrostatistics Special Interest Group, International Statistical Institute (ISI).
- Commission B3 Astroinformatics and Astrostatistics, International Astronomical Union (IAU).
- Astrostatistics Interest Group, American Statistical Association (ASA).
- Working Group on Astroinformatics and Astrostatistics, American Astronomical Society (AAS).
- Informatics & Statistics Science Collaboration, Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST).

https://asaip.psu.edu/

# Useful References


- Siemiginowska, A., G. Eadie, I. Czekala, E. Feigelson, E. B. Ford, V. Kashyap, M. Kuhn, et al. “Astro2020 Science White Paper: The Next Decade of Astroinformatics and Astrostatistics.” arXiv, March 15, 2019. http://arxiv.org/abs/1903.06796.
- Eadie, Gwendolyn, Arash Bahramian, Pauline Barmby, Radu Craiu, Derek Bingham, Renée Hložek, J. J. Kavelaars, et al. “LRP2020: Astrostatistics in Canada,” October 21, 2019. https://doi.org/10.5281/zenodo.3756019.
- Feigelson, Eric D., Rafael S. de Souza, Emille E.O. Ishida, and Gutti Jogesh Babu. “Twenty-First-Century Statistical and Computational Challenges in Astrophysics.” *Annual Review of Statistics and Its Application* 8, no. 1 (2021): 493–517. https://doi.org/10.1146/annurev-statistics-042720-112045.
- Tak, Hyungsuk, Yang Chen, Vinay L. Kashyap, Kaisey S. Mandel, Xiao-Li Meng, Aneta Siemiginowska, and David A. van Dyk. “Six Maxims of Statistical Acumen for Astronomical Data Analysis.” arXiv, August 28, 2024. https://doi.org/10.48550/arXiv.2408.16179.

Don’t forget that statisticians are the free-est of all scientists, they can work on anything.

David R. Brillinger

Image Source: ICRAR