Ben Nelson
Senior Data Scientist at Chan Zuckerberg Initiative
"Evaluating the impact of open-source bioimage analysis software"
The Chan Zuckerberg Initiative is supporting the development of imaging software in the cell biology space. "napari" is a fast, interactive, multi-dimensional image viewer for Python, and the "napari hub" is a central tool for finding and installing community-built plugins to solve the user's analysis needs. These tools have also seen applications in atmospheric science, geophysical science, 3d printing, and even astronomy. In this talk, I will discuss the current vision for napari and my role in gathering data to better understand its usage, development, and discourse within the larger bioimage tech ecosystem.

Betsy Barton
Founder & CEO at Infiniscape
"How do we fix social media?"
In recent years, social media has been accurately blamed for many societal problems including phone addiction, poor body image, depression, political polarization, rampant misinformation, the growth of anti-science movements, and many others. As long as social media newsfeeds are controlled by algorithms that optimize primarily for "engagement," we will continue to see widespread damage. I argue that a new era of transparency is the best solution. If consumers become knowledgeable about the effects of algorithmic approaches and demand different solutions from social media, social media companies will be forced to deliver a better product. I will encourage discussion of the key strategies that will lead to this revolution.

Brendan Wells
Senior Data Scientist at Samba TV
"Project Management for Academics"
Think back to your last project: did you meet your goals? Were you on time? Did you have crunch time in the end, or was it smooth? As academics, nearly everything we do is a project, but rarely are we taught how to handle them efficiently. That subject is known as Project Management, and it is one of the most valuable disciplines for academics but one of the least frequently trained. Every paper, presentation, research objective, and event could be considered a formal project: a unique, finite-duration goal. Without some consideration, it's easy to make simple oversights that cause you to fail your objective, take too long, or spend too much. Alternately, it's also easy to do a little bit of planning, avoid those problems, and smooth the way to success. This talk will summarize the basics of project management, including the most important concepts, a few practical planning approaches, and common failure-and-recovery patterns. You should leave the talk with strategies you can directly apply to your work today.

Christopher Bochenek
Data Scientist at Google
"Job Searching as a Strategy Game"
In this talk, I will give a detailed overview of my recent job search as a new astrophysics PhD. We will discuss what a typical interview process looks like and the core skills required for interviewing for data scientist roles. In addition, I will present a formalism for how to think about job searching as a game to maximize chances of success as well as give strategies that worked and didn't work for me.

Deep Chatterjee
Illinois Survey Science Fellow at the Center for Astrophysical Surveys, NCSA
"Multi-messenger astronomy - a data-science problem"
Time-domain and gravitational-wave (GW) astronomy have gone through a revolution in the last decade - both fields went from a few or no discoveries to routine, even profuse, discoveries. These two previously disjoint fields came together when the electromagnetic (EM) counterpart of a binary neutron star merger was discovered in 2017. This was an unprecedented effort at a global scale. But this has been the only EMGW success story; routine observations using multiple probes is still a problem to be solved. The heterogeneous nature of data from multiple surveys, automated filtering techniques, near-realtime inference needs, and rapid communication between facilities makes this a blend of big data techniques. I have worked on several aspects of multi-messenger astronomy, both from the EM and GW side. Here, I will talk about machine learning applications and data science techniques that are used today, some of which I developed. I will also talk about the current cyber-infrastructure efforts that are underway to enable multi-messenger astronomy in the future.

Eva Noyola
Sr. Data Analyst at Zen Business
"The value of data analysis (not everything is about ML)"
As part of the Data Science group in my company (a startup), I use my data analysis skills much more than any Machine Learning knowledge I have. I feel like too much emphasis is placed on the ML portion of being a data scientist and not enough is placed on the pure data analysis side. I'll try to review the specific data analysis skills that are particularly useful for working in the industry.

Genevieve Graves
CEO at Eye0, Inc.
"State-of-the-Art Natural Language Processing: Teaching Artificial Intelligences to talk, think, and even crack jokes"
The last few years have seen enormous progress in Natural Language Processing (NLP)—the data science sub-field of teaching machines to produce and comprehend human language. The ability to communicate directly with people, in real-time and in ordinary human language, is a critical component of making AI broadly useful across a wide range of industries. It is also likely a prerequisite for true machine intelligence—certainly at least for an intelligence that we can hope to understand and interact with.

In this talk, I will introduce some of the core ideas and tools of NLP, with a focus on recently developed "Transformer Models". Transformer Models make it possible for individual data scientists and small teams to leverage large, powerful NLP models that have been trained (at huge expense of compute time, dollars, and energy) by major institutions. With just a small amount of "fine-tuning" on niche datasets, it is possible to leverage state-of-the-art NLP to solve your specific problem. We'll see what happens when we get a computer to tell jokes, how much we can improve the models with fine-tuning, and some of the major pitfalls that can happen along the way. I will include access to Python code and Jupyter notebooks that implement these Transformer Models in Pytorch, so that you can play around with them yourselves later. In the talk, however, we'll keep discussion at a non-technical level to focus on the main ideas, themes, and challenges in teaching computers to talk.

Jessica Kirkpatrick
Senior Data Science Manager at thredUP
"Search and Taxonomy: Surfacing results when no two items are the same"
thredUP is the world's largest consignment marketplace. One of our big challenges is surfacing the right items to our users when every item is unique. I will be discussing how we approach search and taxonomy to address this problem.

John Franklin Crenshaw
PhD student at the DIRAC Institute at the University of Washington
"Deep Generative Modeling of Astronomical Data with Normalizing Flows"
Generative models are a powerful tool for astronomical data analysis, including simulation, data augmentation, and posterior inference. Recent advances in Machine Learning have enabled Deep Generative Models (DGMs) to learn high fidelity, scalable models of complex, high-dimensional data sets. Normalizing Flows (NFs) are particularly powerful as they are easier to train, do not suffer from mode collapse, and provide efficient sampling and exact likelihood inference. In this talk, I will introduce the basic theory of NFs, including their strengths and weaknesses compared to other popular DGMs. I will then show examples of how the Dark Energy Science Collaboration (DESC) is applying NFs in cosmology research, including forward modeling photo-z systematics and generating realistic host galaxies for simulated SN surveys. Finally, I will introduce PZFlow, a jit- and GPU-enabled python package for powerful, out-of-the-box modeling of any tabular data.

Taka Tanaka
Head of Data & Managing Director at Roc360
"Novel Data Challenges in Residential Real Estate"
Roc360 is a financial firm specializing in residential real estate investments. We are a market leader with a scientific, data-driven philosophy. The projects that our data science team works on includes lead generation, property valuation, owner intent modeling, risk prediction, computer vision, network graphs, and business intelligence reporting. In this talk, I will share the wide range of (largely unsolved-at-scale!) modeling challenges in residential real estate, and discuss the variety of data types and data sources we invoke in tackling them.