School of Data 2019 Schedule

Schedule updated as speakers are confirmed by Hunter Owens, Vyki Englert, and Emily Forscher.

Website http://schoolofdata.la

Venue

Southern California Association of Governments Office @ Wilshire Grand Center
900 Wilshire Blvd., Ste. 1700 Los Angeles, CA, 90017 (17th Floor)
Enter closest to Figueroa Street on the south east side of the building. Inform the receptionist that you are there for the event on the 17th floor. Here is a link to a video developed by SCAG on how to get to the building: https://vimeo.com/252771657

Transportation

- **Public Transportation** We strongly recommend using public transportation. SCAG's offices are across the street from 7th Street / Metro Center Station. SCAG is accessible via Metro Rail from the 7th and Metro Station, by all Metrolink Service to Union Station, or by DASH Routes A, B, E, and F. Bus Service via MTA, Foothill, Santa Monica, and Orange County transit is available downtown.
- **Cycling** There are bike racks on the north east side of the building along Wilshire Blvd. There is also a mobility hub on 7th street, with bicycle lockers for rent.
- **Driving / Parking**: We will not reimburse or validate parking. Please find below a list of affordable parking lots and garages near 900 Wilshire Boulevard in Downtown Los Angeles:
    - § 725 Grand Ave. *Daily maximum: Mon.-Fri. $12*
    - § 718 S. Figueroa St. *Daily maximum: $15*
    - § 724 S. Figueroa St. *Daily maximum: $15*
    - § 818 W. 7th St. Garage: (Entrance at 723 S Flower St.) *Daily maximum: $15*
    - § 725 S. Flower St. *Daily maximum: $16*
    - § 737 S. Flower St. *Daily maximum: $13*
    - § 757 S. Flower St. *Daily maximum: $16*

Audiences

- **Analyst** - you use excel to deliver reports to management. Explore how to automate your work, using tools and workflows such as Shiny Dashboard, Tableau, and Power BI. Taught in R, attendees will practice using R to generate reproducible research and reporting on their desktop.
- **Systems**: you have large databases, how might you support analyses on top of you data? Explore how to plan for future analyses and how to build good data systems. Attendees may manage data in a SQL Data Warehouse or ArcGIS Server, may have to deal with new large data sources or manage data and data products at scale. In this track, most workshops will employ Python.

- **Manager, Electeds:** How do you build and resource a team that uses data? How and why to hiring for data skills? How to deliver data projects? How do you budget for large data projects and initiatives? Do I need to hire, can't I just buy this platform? Programming in this track will involve analog props, no laptops or coding required.

Yellow Highlight = need speakers
Pink Highlight = speaker request sent
Green Highlight = speaker confirmed

Day 1: Thursday January 31st, 2019			
	Analyst Track *R focused* *Regional Council Room*	**Systems Track** *python focused* *Policy B Meeting Room*	**Manager/Electeds Track** *no code, analog track* *Policy A Meeting Room*
Breakfast & Registration 8:00-9:00 am	--- Breakfast & Registration --- Catering by Noah's New York Bagels		
Welcome Remarks 9:00 - 9:15 am	LA Controller Ron Galperin emcee: Jon Neill, County of Los Angeles 9:00 am - Jon Neill: Welcome, thank yous, and introduce LA Controller Ron Galperin (5 minutes) 9:05 am - Ron Galperin greets participants (5 mins) 9:10 am - Jon: Logistics overview (first session locations, breaks, bathrooms, any questions ask D&D team) (3 mins) 9:13 am - Everyone goes to their first session		
Morning 1 9:15 -10:45 am	Introduction to Programming in R *Kevin Kane, SCAG* *NEED TAs* *--- break ---*	How to Make Your Future Data Scientist Love You: Collecting Data Right the First Time Sasha Laundy, Warby Parker Hunter Owens, City of Los Angeles	-- no class / room unavailable --

Break 10:45 - 11:00 am	*3 HOUR SESSION*	*--- break ---*	
Morning 2 11:00am - 12:30pm		Better Dev Environments for Free Using Containers Kegan Maher, City of Santa Monica TA:	-- no class / room unavailable --
Lunch 12:30 - 1:00 pm	Catering by Bánh Mì Chè Cali Options: **BBQ Pork Bánh Mí**	**Lemongrass Chicken Báhn Mí**	**Lemongrass Tofu Báhn Mí**
Remarks 1:00 - 1:05 pm	Dave Wesolik, Internal Services Department, Los Angeles County		
Keynote *1:05 - 1:45pm*	*Leah Bannon, USDS / VA.gov* *founder of Tech, Rebalanced* *Theme: Data and Delivery in the Government*		
Break 1:45 - 2:00 pm	*--- break ---*		
Afternoon 1 2:00 - 3:30 pm	Geospatial Data: Intro to Data Types, QGIS, and Exploring Census Data in R *Andy Rutkowski, USC* *Kate Vavra-Musser, USC*	To and from the ESRIverse, using Geospatial Data inside and outside of the ESRIverse (including Python, ArcGIS, QGIS, Spatial SQL in PostGIS) *John Gravois, Esri* *Alfredo Novoa, UCLA*	Demystifying Data Science 101 Hunter Owens, City of Los Angeles Kegan Maher, City of Santa Monica
Break 3:30 - 3:45 pm	*--- break ---*		

Afternoon 2 3:45 - 5:15 pm	Sharing and Versioning Your Work with Github, Git and the Command Line (GitHub Talk) (description) Nathan Danielsen, Farmers Insurance Patricia Realini, GitHub Kegan Maher (advisor)	Consuming and using APIs, Building ETL Pipelines (description) Dave Marulli and Christopher Tull, California Data Collaborative / ARGO	Hiring Data Skills (Video) Lead: Maksim Pecherskiy, City of San Diego Phillip LeClair, City of Pasadena Behrang Abadi, City of Santa Monica

Day 2: Friday February 1, 2019			
	Analyst Track *R focused* *Regional Council Room*	**Systems Track** *python focused* *Policy B Meeting Room*	**Manager/Electeds Track** *no code, analog track* *Policy A Meeting Room*
Breakfast & Registration 8:00-9:00 am	--- Breakfast & Registration --- Catering by Noah's New York Bagels		
Welcome Remarks 9:00-9:15 am	TBD, SCAG Kimberly Clark, SCAG - Future Cities Framework emcee: Jon Neill, County of Los Angeles 9:00 am - Jon Neill: Welcome, thank yous, and introduce Kimberly Clark (5 minutes) 9:05 am - Kimberly Clark greets participants (5 mins) 9:10 am - Jon: Logistics overview (first session locations, breaks, bathrooms, any questions ask D&D team) (3 mins) 9:13 am - Everyone goes to their first session		
Morning 1 9:15 -10:45 am	Data Cleaning Talk *Jon Schleuss, LA Times*	Databases, Dashboards & Death to Paper Mateo Clarke, City of Austin	Dashboard in 90 min Juan Vasquez, City of Los Angeles

		John Clary, City of Austin	
Break 10:45 - 11:00 am	*--- break ---*		
Morning 2 11:00am - 12:30pm	Reproducible Research and Reporting in R *Lauren Steely, MWD* *Robert Mitchell, Skid Row Housing Trust*	Data Lakes + Data Warehouse and Analytics Datastores Lead: Jon Neill TA: Maksim Pecherskiy, City of San Diego	Governance and Stewardship of Public Data (A survey of case studies) *Mark Greninger, LA County* Toki Sadralodabai, Department of Health Services Drew Atwater, MNWD Moderator: *Taylor Campbell, LA Controller's Office*
Lunch 12:30 - 1:00 pm	Catering by Chinese Laundry **Veggie Super Bowl 素三杯 (Vegan)** Seared king oyster mushrooms and eggplant dry braised with ginger and garlic in dark soy, sesame oil, and rice wine, served with braised beans, sauteed leafy greens, and fried onions Or **Grandma's Hong Shao Pork Belly Bowl 红烧肉** Braised pork belly with cucumber, pickled onions, scallions and cilantro		
Keynote Panel 1:00 - 1:45 pm	Intro: Drew Atwater, Moulton Niguel Water District Theme: Data is Policy Andrew Schrock (Moderator), Marcel Porras, City of Los Angeles, Dr Veronica Barassi, Goldsmiths, University of London Alma Castro, Data Justice Lab / Long Beach		
Lightning Talks 1:45 - 2:00 pm	The "California Model" for Data Collaboratives - Patrick Atwater, CaDC/ARGO DocNow - Andy Rutkowski CalStreetLookup - John Hickok, County of Los Angeles		

Afternoon 1 2:00 - 3:30 pm	Automating reports with Power BI so you can do more with your time. *Lead: Jessica Lam City of Los Angeles Bureau of Contract Administration* *TA: Eric Lai, City of Los Angeles, Bureau of Engineering*	Predictive Models in Production Medium post *Will High, Joymode + DataKind* Nishchal Chaudry, City of Long Beach	Critical Ambiguity: On the Impact of Bias in Data Patricia Realini, GitHub TA: Leah Bannon, USDS
Break 3:30 - 3:45 pm	*--- break ---*		
Afternoon 2 3:45 - 5:15 pm	Shiny: Building interactive, shareable web apps in R *Lauren Steely, MWD* *Robert Mitchell, Skid Row Housing Trust*	Simple Machine Learning / Why Model *Ian Rose, Project Jupyter* ***https://tinyurl.com/la-ml-demo*** *NEED TA*	Data Viz 101 with Felt Cutouts (Based on WPRDC Chart workshop) Alex Pudlin, City of Los Angeles Umi Hsu, City of Los Angeles

Session Descriptions

Day 1 Keynote

Title: Data and Delivery in the Government

Speaker(s)/Instructor(s): Leah Bannon

Description:

Leah has worked for over a dozen agencies as a federal employee, contractor, sub, 18Fer, and US Digital Service product manager. She is also a community organizer: she lead a Code for America brigade of volunteer civic hackers in DC -- and she founded Tech, Rebalanced, a day for folks of marginalized and underrepresented genders to gather for civic hacking and training in a supportive environment.

She'll talk to us about shipping products that people like --- and surviving --- in government, and then she'll be taking your questions

Analyst Track Sessions

Geospatial Data: Intro to Data Types, QGIS, and Exploring Census Data in R

Learn more about the basics of what makes data spatial. We will walk through basic spatial data types, next you learn how to use QGIS to explore,manage, and visualize data, and lastly you learn how you can use R to find and work with census data.

Please come to this workshop with QGIS (latest version preferred), R, and R Studio installed. Also, please get your census API key here: http://api.census.gov/data/key_signup.html

Download R Notebook - http://bit.ly/zipsod2019

Sharing and Versioning Your Work with Github, Git and the Command Line

Have you heard about Github? Have you wondered if there is a better way to version your work then constantly changing the file name and adding the date to it?

Also wonder why some data scientists and developers always seem to be typing into the command line? What is the benefit of the command line over using clicking through multiple windows?

In this workshop, we will cover:

- What is github and why is it facilitates collaboration and the sharing of work
- The basics of how and why to use the terminal (aka the command line)
- How to use git to version control your work
- How using these tools can increase your impact and effectiveness

Before the workshop, please follow the set-up instructions located here:

https://github.com/ndanielsen/intro-cli-git-github

Automating reports with Power BI, so you can do more with your time.

During this presentation, we will be demonstrating the various functionalities of Microsoft Power BI to create insightful and interactive visualizations. Learn how to create a dashboard using Power BI for any specific business purpose and translate business rules into code using raw data!

Topics we will cover include:

- Dashboard and basic relational database terminology.
- The Systems Development Life Cycle applied to dashboard creation.
- Creation of a dashboard using sample data from the Bureau of Contract Administration.
- Programming using Data Analysis Expressions (DAX) to calculate measures used to solve data analysis problems.

You will need a laptop with a Windows operating system. Please install Microsoft Power BI Desktop, here. Power Bi Desktop is free to use and download. We will be using the Bureau of Contract Administration's sample payroll data to create an interactive report. Download the data as a csv file, here. We will be programming in DAX! Get the code here.

Reproducible Research and Reporting in R

So you've spent hours on a cool data analysis project - how do you share it with your colleagues, your stakeholders, or the public? And how do you ensure other analysts can reproduce what you've done?

In the first part of this class, we'll discuss how to organize and document your data projects so others can understand and recreate what you've done. We'll address good file management, data pipelines, and documentation practices. In the second part, we'll discuss ways to turn your analyses into reports, web pages, and dashboards using open-source R tools such as Rmarkdown, shinydashboard, and flexdashboard.

Shiny: Building interactive, shareable web apps in R

Reports and dashboards are great, but sometimes you want to let your audience explore the data themselves in a more interactive fashion. Shiny is a tool for building lightweight, interactive web apps that is relatively easy to learn. In this interactive course, we'll build several small apps together and look at some real-world applications.

System Track Sessions

Better Dev Environments for Free Using Containers

Do you write code or process data in a team environment? Do you ever have to test new features or bug fixes before deploying them to a production system? Are you fed up with your government mandated workstation OS and software packages, because it all just holds you back from the latest bleeding-edge development goodness? In this session, we'll skip procurement and central IT and ramp up quickly into ETLing and analyzing data using python, SQL, and your OS of (mandated) choice.

Consuming and using APIs, Building ETL Pipelines

This session discusses techniques to help tame the chaos of public sector data integration projects. Apache Airflow is introduced as a Python workflow management tool that allows ETL processes (including pulling from APIs) to be coordinated and scheduled using code. Participants will get practice "thinking in DAGs" as they design their own workflows for data integration.

Predictive Models in Production

Follow along in this Medium post that was written for this session. We will get through as much of it as we can. We're likely to

- set up the local Docker environment,
- set up the remote Docker image repository,
- get the AWS Lambda function running, if we're plucky, and
- *not* get the AWS Batch model trainer system set up because it requires you to ask AWS to increase their limits, which takes a day or so.

Speaker Bios

Alphabetical by First Name

Andy Rutkowski

Visualization Specialist, University of Southern California Libraries

Andy Rutkowski is a currently a Visualization Specialist at USC Libraries. He has worked in academic libraries since 2002, specializing in Government Information and Documents (NYU), GIS and Geospatial Librarianship (USC and UCLA), and now more broadly visualization. He is a co-director on an IMLS (Institute of Museum and Library Services) grant, Visualizing the Future Symposia: A National Forum on Data Visualization. Twitter: @andy_rutkowski

Christopher Tull

Christopher Tull is the lead data scientist for ARGO's water data work where he designs and builds data infrastructure to process, store, and analyze metered water use data from water utilities across California. Prior to that he worked in the Urban Intelligence Lab at New York University where he developed the official energy benchmarking analysis website for the City of New York and forecast energy use for all 1.1 million buildings in NYC. He has an M.S. in Urban Informatics from New York University and a B.S. in Computer Science and Mathematics from CSU Channel Islands. Twitter: @ctullito, LinkedIn: ctull

David Marulli

David Marulli is a data scientist working across the disciplines of data engineering, analysis, and design. With the California Data Collaborative, he has led the development of the Statewide Efficiency Explorer tool that delivered the world's first estimates of Governor Brown's water efficiency framework. David is leading other ongoing statewide data integration efforts including a new parcel-level turf conversion savings calculator which will be available for all Californians. Prior to working with ARGO, his work was sponsored by NYU's Rudin Center for Transportation Policy and Management, and the Urban Intelligence Lab. He holds a Masters in Urban Informatics from NYU and Bachelor degrees in Physics and Philosophy from UW-Madison.

Hunter Owens

Hunter Owens is the Senior Data Scientist at the City of Los Angeles Information Technology Agency. Prior to joining the City, he worked for the University of Chicago's Center for Data Science and Public Policy, KIPP NJ, and Obama for America. You can find on Github and Twitter.

Jessica Lam

Jessica Lam is the data analyst for the Bureau of Contract Administration, Department of Public Works with the City of Los Angeles. She analyzes current workflows for ways automation can help deliver reliable and up to date stats on local workforce development for the City's Public Works construction projects. Her work involves creating interactive reports using Power Bi, taking into account business rules and database constraints; and enjoys empowering management decisions with data analytics. Check out one of her reports here, and learn what Power Bi is all about here!

Jon Neill

With over 30 years of IT experience at the County of Los Angeles, Jon Neill has been a visionary and technology leader at both the Internal Services Department (ISD) and the Auditor-Controller. Jon is currently a Senior IT Specialist in ISD's Customer Applications Branch leading the Data Analytics and Business Intelligence team. He is spearheading ISD's transformation on how technology is delivered to its customers including Big Data, Agile, and DevOps programs, pushing the County towards becoming a modern digital organization.

Jon is also involved in civic technology outreach bringing government technology workers together and is a founding organizer of Data and Donuts. Jon is an avid golfer, curler, and computer gamer.

Kate Vavra-Musser

Doctoral Candidate, University of Southern California Spatial Sciences Institute

Kate Vavra-Musser is a doctoral student in the Population, Health, and Place PhD program with the University of Southern California's Spatial Sciences Institute. Topics she enjoys studying include integrated human-environment systems, environmental health and justice, and climate change. She is passionate about all things geospatial and data-driven. You can find Kate on LinkedIn and Twitter.

Kegan Maher

Kegan Maher has been working with the City of Santa Monica since before Obama's first term, and he's still finding new things to hack on. In his current role as the City's Data Officer, Kegan oversees Santa Monica's open data and open source publications, works on regional transit and mobility issues, staff education and skill-building, and organizes Hack for LA's Westside Hack Night on Mondays in Santa Monica. GitHub: @thekaveman

Lauren Steely

Lauren Steely is a water resources analyst and data scientist with a passion for turning data into policy. Her work focuses on the science and economics of urban and agricultural water use in the western U.S. She received an MS from UC Santa Barbara, where she participated in the Sustainable Water Markets program at the Bren School of Environmental Science and Management. She tweets at @MadreDeZanjas.

Leah Bannon

Leah has worked for over a dozen agencies as a federal employee, contractor, sub, 18Fer, and US Digital Service product manager. She is also a community organizer: she lead a Code for America brigade of volunteer civic hackers in DC -- and she founded Tech, Rebalanced, a day for folks of marginalized and underrepresented genders to gather for civic hacking and training in a supportive environment.

Marcel Porras

Marcel Porras is the Director of the City of Los Angeles' Department of Transportation Shared Mobility Group. The Shared Mobility group oversees taxis, bike share, car share, rideshare, and autonomous and connected vehicles. Previously Porras served as Mayor Garcetti's Director of Transportation. LinkedIn

Mark Greninger

Chief Data Officer, County of Los Angeles

Since 2016, as the County's Chief Data Officer, Mark's goal is to facilitate countywide information sharing and management, support data driven decision making, and provide a 360 degree view of county residents. This will help ensure the County provides the right service at the right time to ensure the best outcomes for its residents.

Nathan Danielsen

Senior Backend Developer, Farmers Insurance

Nathan is a core contributor to Yellowbrick, a python open source machine learning visualization library. Nathan is also a Senior Backend Developer for Mobile Apps at Farmers Insurance. Previously, he worked as a full stack developer at the Atlantic Media company. In his spare time,

Nathan enjoys gardening, enjoying the southern California climate and is training for his first Ironman triathlon. He served in the Peace Corps in Senegal after graduating from UC Berkeley.

Umi Hsu

Digital Strategist/Senior Project Coordinator, City of Los Angeles Department of Cultural Affairs

Twitter: wfumihsu LinkedIn: https://www.linkedin.com/in/w-f-umi-hsu-phd-21799b11/

Umi Hsu is a researcher, strategist, and educator who engages with hybrid research and organizing agendas for equality in arts, technology, and civic participation. A former ACLS Public Fellow, Hsu currently works as the digital strategist of the City of Los Angeles Department of Cultural Affairs, providing research and strategy to redesign the department's data and knowledge architecture. They are also the founder of Lab at DCA, a city staff innovation incubator. Hsu has served on the advisory committee for Arts for LA, Center for Cultural Innovation, California Community Foundation's Social Change Data Commons, Cultural Research Network, and the Society of Ethnomusicology Council. They also lead two community-driven arts collectives LA Listens and Movable Parts. Since 2007, Hsu has taught at University of Virginia, Occidental College, and Art Center College of Design.

William High

Head of Data Science, Joymode
Data Ambassador, DataKind

Will heads analytics, machine learning and experimentation at DTLA startup Joymode. He spent four years at Netflix LA as a Senior Data Scientist building machine learning models around movies and TV shows. Since about 2014 Will has also acted as technical lead on DataKind projects once or twice a year, including water utility and homelessness problems in Los Angeles. He holds a PhD in astrophysics from Harvard and did research at Berkeley, Caltech and UChicago prior to joining the private sector. In total Will's got 15 years of experience in scientific experimentation, analysis and systems. Feel free to connect on Twitter @WillHighSci and on LinkedIn.