

DataFest at CSU, Chico 2019 - Participant Info & Guidelines

<https://chicodatafest.netlify.com/>

Note: This document may be updated as the event approaches; any major updates will be clearly marked.

This is a long document, please take a moment to read over it carefully.

Location

DataFest 2019 @ CSU, Chico will start with registration and end with awards ceremony in Selvester's Cafe. In between, participants will work during the day and night at Selvester's Cafe. See the schedule for more detailed information for important events in between and their locations.

Directions & Maps

- Selvester's Cafe
 - Is in the center of campus, along the creek
 - For some reason it's labelled Selvester's Cafe-By-The-Creek
 - On a map: <https://www.csuchico.edu/maps/campus/?id=1193#!ct/28514?mc/39.72979688502985,-121.84459609910849?z/18?lvl/0>

Transportation & Parking

- There is metered parking (free on weekends) on Citrus Ave and Legion Ave.
- There is additional free street parking near the Gateway Science Museum.
- Please don't park in the Bidwell Mansion parking lot or in the residential areas.
- CSU, Chico offers some further advice about parking.
- We ask that you carpool with teammates as parking on campus can be tricky.
- Here is a map of some parking spots on campus.

Who is at ASA DataFest @ Chico State 2019?

- 46 participants in 11 teams!
- 8 consultants
- 4 judges

You should receive an email from the DataFest organizers with your team information, this document as a PDF, and information on how to join our Slack workspace. Contact us if you have not received this information.

Schedule

The schedule is at <https://chicodatafest.netlify.com/schedule/>.

Registration opens at 5:00pm on Friday at Selvester's Cafe.

You are of course free to come and go as you please throughout the event, but here are the times all team members should plan to be on premises:

- Friday sometime between 5-5:30pm - registration
- Friday 5:30pm - kickoff
- Saturday afternoon - group photo
- Sunday 2pm - judging and awards ceremony

Consultants will be available for help until midnight; you can work as late/early as you like. We recommend checking out the Consultant schedule to plan out your weekend.

Security & Access

Campus security will open the doors promptly at 8am Saturday and Sunday morning, and come to lock the doors at Midnight Friday & Saturday. It is not advised to leave materials in the building although it will be locked.

Teams who are spotted working in areas other than Sylvesters or after midnight will be disqualified from the competition.

SOCIAL MEDIA

Did you know that, this year, DataFest will be held at over 30 universities around the world? DataFest takes place during a six-week window in the Spring, and different universities hold their event at different times.

An essential ingredient of Datafest is that the data be a surprise. There are several reasons for this. One is that it ensures that all teams start on an equal footing. Another is that, even if you research the organization that donated the data, you might still be totally unprepared for the context of the data (and maybe less prepared than had you known nothing!) Finally, it is simply more fun.

You are encouraged to share your excitement on social media and thank our sponsors. The official Twitter hashtag is #datafest19. However...

PLEASE KEEP THE SECRET UNTIL MAY 4, 7pm. Teams that leak the name or any information about the identity of the data donor will be disqualified!

Please do not post pictures that reveal any charts/graphs/tables/summaries/models that might reveal the data donor.

Computing and supplies

We recommend that every member of the team bring a laptop, if possible. You might find it helpful to have a mix of PCs and Macs, since they have different strengths.

We recommend that you make sure beforehand that the software you will be using throughout the weekend is properly installed and running on your computer. You will be working with a large dataset so make sure that you have the space for it on your hard drive.

You might want to bring some favorite statistical or computational reference books, if you have them, or bookmark some pages that you routinely refer to.

We will provide meals, snacks, and munchies. Feel free to bring anything additional you might want.

Cloud computing resources for DataFest

RStudio is hosted on the cloud at <https://rstudio.cloud/>

You can use a Python instance from Google Colab

Both of these options are accessible to anyone with a gmail account.

Data

At the end of the kickoff presentation each team should send one member to the registration desk at Selvester's Cafe and check out a USB stick containing the data. When you're done downloading the data off it please return it to the registration desk so that another team can use it.

Large data advice

The dataset you will be working with is quite large. If you type a variable name to view it, it will take a while to display. Therefore, remember these R commands: `head()`, `tail()`, `str()`.

We strongly recommend you create a small data set that you can use to test things on. Then, if it works out, you can apply your procedure to the large dataset. Some procedures can take a frustratingly long time to run on large data sets, and so it will be comforting to know that your procedure works (because you tested it on a smaller data set) while you wait. We recommend taking a random sample of rows from the original data set, but there might be other approaches you find useful.

Presentations, judging, and awards

Presentations

Each team will have 4 minutes + 1 minute Q&A to present their findings to the judges. That's exactly 4 minutes, not 4 minutes and a few additional seconds. Each team will be allowed at most two slides. Two! So at some point Saturday night or Sunday morning, you might want to set aside time to think about what you want the judges to know. The 4 minute presentation and 1 minute Q&A time limits will be strictly enforced. All team members must be present for the presentation, but not all team members need to actually speak (given the time limitation).

Optional

Along with your presentation you are also allowed to turn in a one-page write-up of your project. You can think about this as the text of your presentation. The judges can refer to these during deliberation. You will upload these documents along with your presentation. Only 1 page will be printed.

Submitting your presentation

At 1pm on Sunday all work must stop and you must upload your presentation and your optional write up to the submission website linked at the top of this page (available starting Sunday AM). If you are having technical difficulty, you can ask a consultant for help. Consultants will be around to help.

Teams who fail to upload their presentations and write-ups by 1:0pm on Sunday will not be eligible to have their presentations judged.

File naming

The files you're submitting must be named in the following manner:

- [Team Name] - Presentation
- [Team Name] - Writeup

Allowed file formats

- We very strongly recommend using PDF, Keynote, or PowerPoint.
- If using a web-based tool like GoogleDocs or Prezi, please export to PDF and upload the PDF as your submission.

Note that you will not have time to log on/off to your account before your presentation. We don't want to restrict your creativity but it is your responsibility to make sure that your presentation works seamlessly before the judging session begins.

JUDGING

The Judges will convene in a side room to deliberate and rank their nominations. Three of these will be selected for the award categories listed below. The judges also have the option to name a fourth winner as Judges' Pick.

All are welcomed to the presentations and award ceremony.

Awards

Awards will be given in three categories:

- Best Insight
- Best Use of Outside Data
- Best Visualization

These are listed in no particular order.

The judges also have the option to name a fourth winner as Judges' Pick.

Award ceremony will take place in Selvester's Cafe.

Winners will receive medals and books as well as one-year student memberships to the American Statistical Association. See <http://www.amstat.org/membership/> for membership benefits.

Raffle prizes

Throughout the event we will be giving out raffle prizes. Announcements for these will be shared on social media. Follow these channels to get a chance to win one of these sweet prizes! Winning will also require that you are on premises at the time a prize is announced.

Recruiting

DataFest is a great recruiting opportunity for many employers, and surely they won't miss it!

Many of our sponsors are attending the event so you can find out more about them.

Most of our consultants are coming from companies who are recruiting or at a minimum wanting to meet you, so chat with them, find out what they do, network.

We will collect resumes and share them with some of our sponsors. Participation in the resume book is optional, but highly recommended. You will receive information about this during the event.

Rules

- You can come and go as you please, but all work must be completed on premises.
- Do not use any space other than Sylvesters, or work after midnight. Students found working in other areas or after midnight will be disqualified from the competition.
- Do not share the name of the data source publicly or on social media before May 4th. There are many other upcoming DataFests around the country and we want to make sure the dataset remains a surprise for them.
- Before your team gets the data all members must have signed Non-Disclosure agreement on file at the registration desk. You can freely share your results, presentations, findings, etc. as part of your digital portfolio, however you are not allowed to share the raw data with anyone outside of DataFest. At the end of DataFest, you must delete all data from thumb drives, hard drives, etc. The data are sensitive.
- As much as possible during the event there will be a friendly consultants present. These are faculty, grad students, or other professionals from our community. Not all will have field specific knowledge on the dataset. They all have different areas of expertise, so if you get stuck on something and one consultant isn't able to help, ask someone else later. Feel free to ask anything. This is not an exam, but a collaborative competition. Do not expect the consultants to write code for you, or do data management, etc. They are there to help point you in the right direction, but you're responsible for getting there on your own.
- PLEASE KEEP THE SECRET UNTIL MAY 4, 7pm. Teams that leak the name or any information about the identify of the data donor will be disqualified! Please do not post pictures that reveal any charts/graphs/tables/summaries/models that might reveal the data donor.
- Please do not tweet, hashtag, Facebook, snapchat, etc about the identify of the donor or the context of the data. This means you should refrain from any hints, explicit or implicit statements that might reveal the context of the data.
- Be careful about github repositories. Make sure yours is invisible to the outside world.