ASA DataFest™
Willamette University

March 31 - April 3, 2022

Calling all data enthusiasts!
All majors and skill levels welcome

Willamette University Data Science is hosting our second annual ASA DataFest. Last year our Datafest brought together 42 undergraduate students (from freshman to senior), mentors, and volunteers from six schools including Willamette University, Reed College, Portland State University; Pacific University, Lewis and Clark College, and Oregon State University. This is a chance to bring together and grow the data science and statistics community in Oregon.
The American Statistical Association has been having DataFest events across the country for eleven years. DataFest is a 48 hour data analysis competition, where undergraduate students work in teams (of 2-5 students) to gain insight on a large, rich, and complex dataset provided by a secret data sponsor.

This is a great opportunity to work with real world data, have fun with friends, and add an awesome experience to your resume.

Kick-off Event at 4pm Thursday, March 31st In-Person at Willamette Salem
(Zoom option also available)
This year our DataFest will be held in a HYBRID format to promote engagement and increase student participation. We are hoping to incorporate in-person team collaboration and mentorship this year!

Student Google Form Sign-up: https://forms.gle/PdRCiJvjLLCEMETG6

Professor/Mentor Google Form Sign-up: https://forms.gle/2keS43geSvwykizY7

Prizes awarded for:
- Best in Show / Best Insights
- Best Data Visualization
- Best Use of External Data

What is ASA DataFest? https://ww2.amstat.org/education/datafest/

Non-disclosure: *Due to the competitive and sensitive nature of these data, please do not leak pictures, slides, materials, or information that provides clues or direct knowledge about the data source until May 3.*

If you have any questions, please contact:
Professor Heather Kitada Smalley : hsmalley@willamette.edu

FAQ:

- **How will we share the data? How will we keep the data secure?**
  Data will be shared with the teams via a Google drive folder starting at 4pm March 31st. Access to this folder will close on Monday April 4th at 12am. This Google Drive folder will contain the dataset as well as helpful resources. Datasets must be deleted from student computers after the competition.

- **Do I need to sign up as a team or can I sign up as an individual?**
  You can sign up as a team or individual. Individuals will be grouped together into teams for the event. Individuals may be grouped with other individuals from different institutions. What a great opportunity to meet other students! Efforts will be made to balance programming skills levels.

- **What tools do you suggest using for virtual collaboration?**
  This DataFest will take place completely online. We will be employing Zoom, which has nice screen sharing and remote access capabilities to work together. Google Drive with Google Docs and/or JamBoards is a good tool for virtual synchronous collaboration. If you wish to write together in LaTeX, Overleaf is a useful tool.
  NOTE: Due to the competitive and sensitive nature of these data, you may not host these data or code in a publicly accessible place like GitHub or Rpubs.

- **Do I need to use R?**
  You can use any programming language or software that you desire! However, help and resources will be offered in R.

- **Do we have to work the entire 48 hours?**
  No, 48 hours is the upper bound of time for this competition. Note that there is a cut-off at midnight each night so that students can rest and take a break. We encourage students to practice a healthy work-life balance and to take care of themselves.

- **When and how should the presentation be submitted?**
  We ask that the presentation be pre-recorded and emailed. This can be done with Zoom while using the screenshare option. You can also feel free to use other tools to accomplish this.

- **How can I help as a graduate student/professor/professional?**
  We are looking for volunteers/mentors to help the students with brainstorming and even a little troubleshooting. Volunteers don't need to be domain experts. Some programming and/or statistics knowledge would be helpful. We are also looking to recruit mentors to make short 5 minute mini-lessons/tutorials that can be pre-recorded or hosted on RPubs.
  If you are interested in helping with this please sign-up on the Google Form:
  https://forms.gle/2keS43geSvwykizY7

Tentative Timeline:

Proposed Schedule:

Thursday (March 31): 4pm - midnight* (8 hours)
- LIVE Kick-off at Willamette Salem (Zoom option also available)
  - Live cast from host
  - Video from data sponsor
- Sprint! / Brainstorming with group
- Google Drive folder with data made available
- Snacks and social time to meet other students

Friday (April 1): 8am - midnight* (16 hours)
- Teamwork time
- Mentor time for sign-up
- Workshops/tutorials available

Saturday (April 2): 8am - midnight* (16 hours)
- Teamwork time
- Mentor time for sign-up
- Workshops/tutorials available
- Finalize findings
- Start working on presentation

Sunday (April 3): 8am - 4pm (8 hours)
- Put together presentation
- Video presentations due at 4pm

Total Challenge Time: 48 hours

Note: Cut-offs at midnight