Visualization

2019-2020 Supercomputing Challenge Kickoff

Introductions

David Janecky – Los Alamos National Laboratory (“retired” scientist)
geologist/geochemist with field, experiment, & modeling experience

Destiny Erickson

Students, Teachers, & Mentors

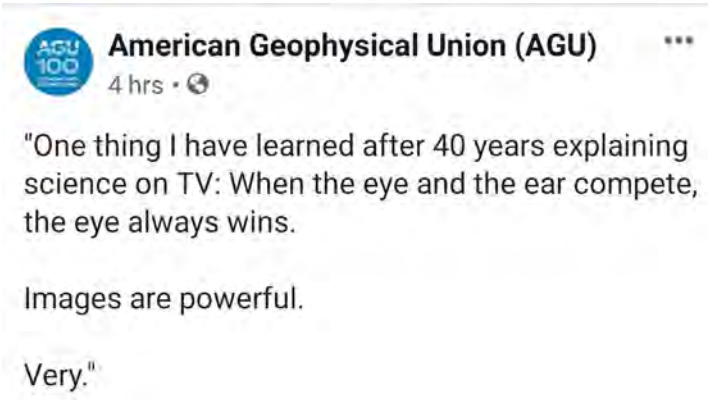
Visualization

communicating your knowledge & ideas

- Data in a figure or chart
- Poster presentation
- Oral presentation visual aids
- Project written report
- Web pages
- Project plan vision & revision(s)

Visualization covers a range of products and scales
from a single figure or chart to the project plan, and project report materials (written report, poster presentation, oral presentation visual aids, website, and more)

A Note about visualization



https://blogs.agu.org/wildwildscience/2019/10/04/ed-Hawkins-climate-stripes-may-be-the-most-important-science-image-so-far-in-this-century/

This twitter post from last week is a parallel statement to the classic “A picture is worth one thousand words”

The modern use of the phrase is generally attributed to Fred R. Barnard. Barnard wrote this phrase in the advertising trade journal *Printers' Ink*, promoting the use of images in advertisements that appeared on the sides of streetcars.[4] The December 8, 1921, issue carries an ad entitled, "One Look is Worth A Thousand Words." Another ad by Barnard appears in the March 10, 1927, issue with the phrase "One Picture Worth Ten Thousand Words", where it is labeled a Chinese proverb. The 1949 *Home Book of Proverbs, Maxims, and Familiar Phrases* quotes Barnard as saying he called it "a Chinese proverb, so that people would take it seriously."[5] Nonetheless, the proverb soon after became popularly attributed to Confucius. The actual Chinese expression "Hearing something a hundred times isn't better than seeing it once" (百闻不如一见, p *bǎi wén bù rú yī jiàn*) is sometimes introduced as an equivalent, as Watts's "One showing is worth a hundred sayings".[6] This was published as early as 1966 discussing persuasion and selling in a book on engineering design.[7]

from https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words

Data Visualization

What makes a good figure or chart?

My experience guidelines are:
- The visualization must be readable, quickly
- Present one key idea in each figure, chart, and slide
- Use colors and distinctive graphic elements (but not too many)

Ask the students to provide some ideas on what makes a good figure or chart

Then show my guidelines

readable quickly

one key idea in each slide – at most one slide per minute in a talk, it takes some time for an audience member to get the concept of each slide (even if some slides have repetitive structure)

colors, symbols, and pictures are easily recognized and distinguished – but caution about using too many

blue/red issues…

Visualization examples from scientific literature
presented in the following slides

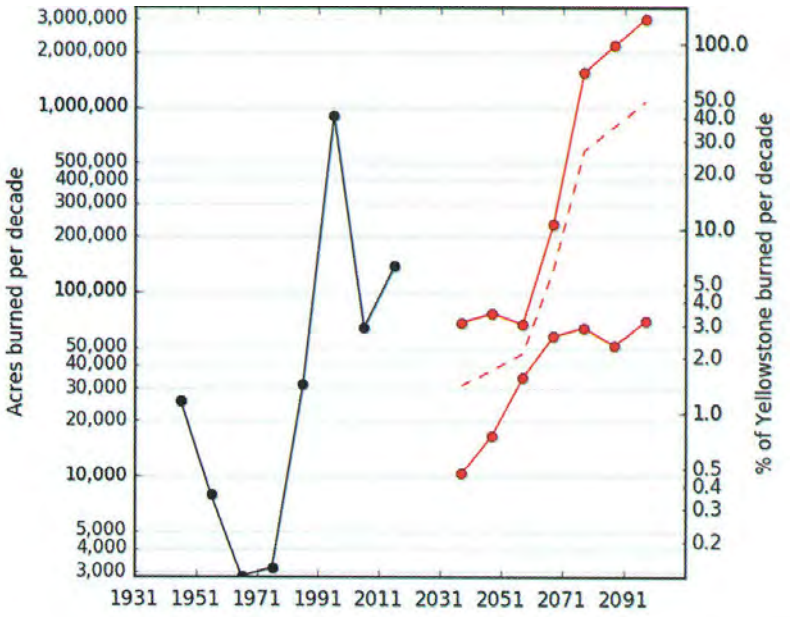
Wildfire & Climate change
from a recent publication on Yellowstone National Park science

Large animal extinction
from a recent paper in Science magazine (American Association for the Advancement of Science)

Big Data
from a recent GSA Today article (Geological Society of America)

Comparison slide examples (bad & good)
from The Oceanography Society

Emphasizing presentation of data from computer simulation modeling



from Tercek (2019) Yellowstone Science 27(1)

Ray et.al. (2019)
Yellowstone Science 27(1)

10/12-13/2019

7

Example visuals (DRJ notes & comments) - The left figure is a very good summary of a research project, including both observed historic data (black) and model forecasts for fire severity. I included the satellite imagery for context and comparison. In a report, these visualizations would be separate with the imagery presented early in the report. In a presentation, the imagery would be presented early and then could be included on this summary slide and used as part of the discussion points.

For the left figure, the axes should be simplified. For the horizontal axis, I find the years used initially confusing and it took me longer to process the scale (loosing some of my attention on the results while I figured it out, a problem when giving a presentation). I would suggest using 50 year increments, and rotating the number to a vertical format. The left vertical axis labeling is overly complex and I would include only major log unit numbers (e.g. 1,000 10,000 100,000 1,000,000). Similarly on the right vertical axis use only major log units (0.1 1.0 10.0 100.0). Normally, I would not include the horizontal lines in the figure, however, in this case they provide good connection between acres and % of Yellowstone. The other aspect to improve readability is to make the tick marks longer on all axes.

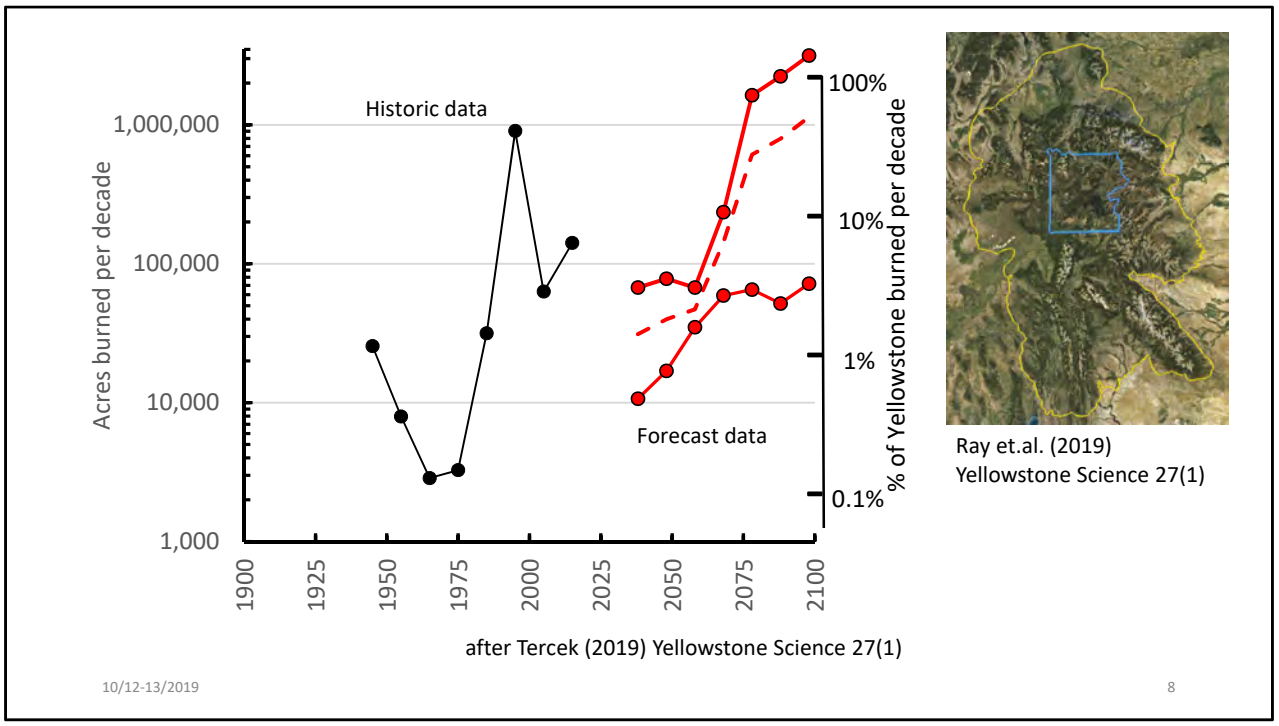
Fig4 caption: Historical and forecast fire severity expressed as acres burned per decade. The black line represents actual historical fire observations. The red lines are future forecast fire sizes. Burn forecasts are based on an emission scenario, which is similar to the

current observed growth in greenhouse gas emissions. The three red lines correspond to the maximum, minimum, and median of the three future data sets that were used as inputs into the elastic net regression. Note the logarithmic scale on the y-axes.

Mike Tercek (2019) Nowcasting & Forecasting Fire Severity in Yellowstone, Yellowstone Science 27(1) 27-33

Fig2 caption: Satellite image of the Greater Yellowstone Ecosystem (GYE showing variations in vegetative productivity (i.e., greenness) across the ecosystem. The GYE boundary is shown in yellow and, for reference, the YNP boundary is shown in blue. Imagery is from the National Aeronautics and Space Administration (NASA).

Andrew M. Ray, David P. Thoma, Kristin L. Legg, David M. Diamond, & Andrew J. Hasen (2019) Assessing the Ecological Health of the Greater Yellowstone Ecosystem. Yellowstone Science 27(1) 18-21



Remade the figure with better axis labeling and structure. I like this example for the Supercomputing Challenge because the main figure combines both real data and a summary of computed simulations (forecast data)

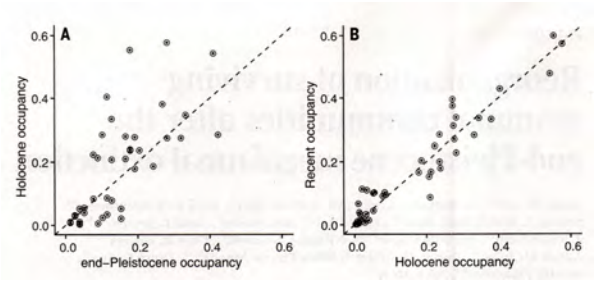
Example visuals (DRJ notes & comments) - The left figure is a very good summary of a research project, including both observed historic data (black) and model forecasts for fire severity. I included the satellite imagery for context and comparison. In a report, these visualizations would be separate with the imagery presented early in the report. In a presentation, the imagery would be presented early and then could be included on this summary slide and used as part of the discussion points.

For the left figure, the axes were simplified. Normally, I would not include the horizontal lines in the figure, however, in this case they provide good connection between acres and % of Yellowstone. The other aspect to improve readability is to make the tick marks longer on all axes.

Fig4 published caption: Historical and forecast fire severity expressed as acres burned per decade. The black line represents actual historical fire observations. The red lines are future forecast fire sizes. Burn forecasts are based on an emission scenario, which is similar to the current observed growth in greenhouse gas emissions. The three red lines correspond to the maximum, minimum, and median of the three future data sets that were

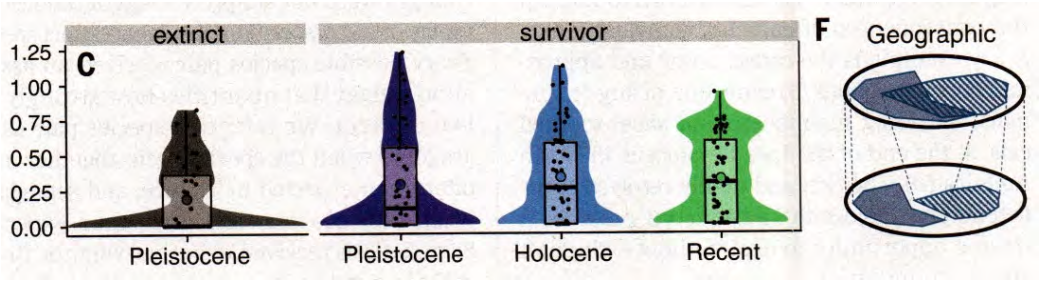
used as inputs into the elastic net regression. Note the logarithmic scale on the y-axes.
Mike Tercek (2019) Nowcasting & Forecasting Fire Severity in Yellowstone, Yellowstone Science 27(1) 27-33

Fig2 published caption: Satellite image of the Greater Yellowstone Ecosystem (GYE showing variations in vegetative productivity (i.e., greenness) across the ecosystem. The GYE boundary is shown in yellow and, for reference, the YNP boundary is shown in blue. Imagery is from the National Aeronautics and Space Administration (NASA).
Andrew M. Ray, David P. Thoma, Kristin L. Legg, David M. Diamond, & Andrew J. Hasen (2019) Assessing the Ecological Health of the Greater Yellowstone Ecosystem. Yellowstone Science 27(1) 18-21



1) Species occurrence & occupancy

Megafauna in North America in Pleistocene
To Holocene and Recent
from Toth *et.al*. (2019)



2) Hypervolume ratios (occurrences : background)

10/12-13/2019

9

Example set – Megafauna in North America in Pleistocene to Holocene and Recent

Add discussion of the readability and complexity of this single slide. In fact, the lower panel was extracted from a much more complicated 6 component figure in the published paper.

A key example of large bodied mammals extinction consequences is the catastrophic and approximately synchronous extinction of large mammals, including mammoths and saber-tooth cats, at the end of the Late Pleistocene in North America. The rich and highly resolved Pleistocene and Holocene fossil record provides a unique opportunity to explore how extinction alters communities.

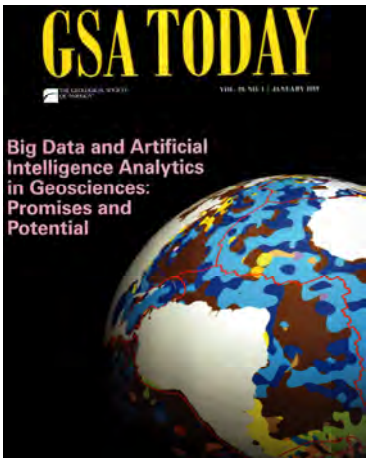
Across the Pleistocene-Holocene transition, common surviving species became even more common and rare species remained the same or became rarer (Figure 1A). There were not substantial changes in occupancy patterns between the Holocene and the Recent (Figure 1B).

Extinction victims had smaller geographic envelopes than survivors in the end-Pleistocene (Figure 2). On average, geographic envelopes of surviving species expanded from end-Pleistocene to the Holocene.

Toth et.al. (2019) Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. Science 365(6459) 1305-1308

Figure 1. Comparison of survivor occupancy across time intervals. (A) End-Pleistocene to Holocene (N=44). (B) Holocene to Recent (N=45). Points are species. The line of unity is shown.

Figure 2. Increases in niche overlap (A to F) Climatic [(A) and (B)] and geographic © envelopes of species are compared to pooled climate envelopes (A) and background envelopes [(B) and ©] in each time interval. In (a), larger ratios correspond with larger riches because niche space expands, as illustrated by oval sizes in (D), in (B) and ©, larger ratios result from proportionately higher fill that causes increased niche overlap [€ and (F)]. In (A) to ©, each shaded distribution sums to an area of 1, circles are means. In (D) to (F), shared polygons represent hypothetical species niches.



Roberto Spina (2019) GSA Today 29(1) 42-43

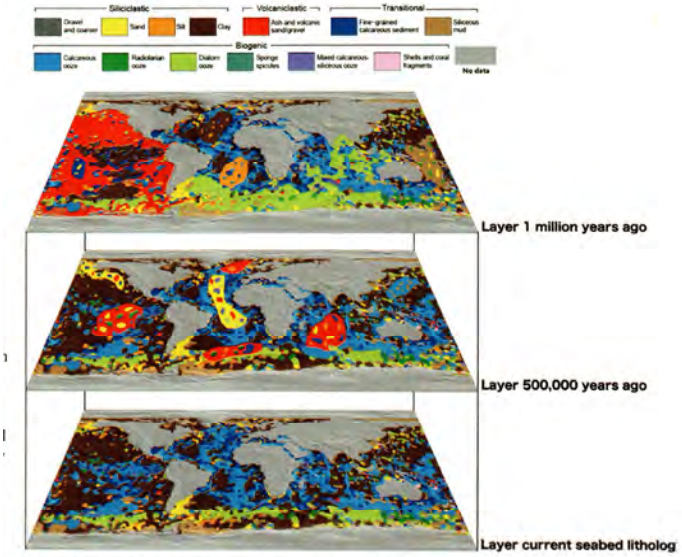


Figure 1. Example of a layered implementation of seabed lithology maps (modified from https://portal.gplates.org).

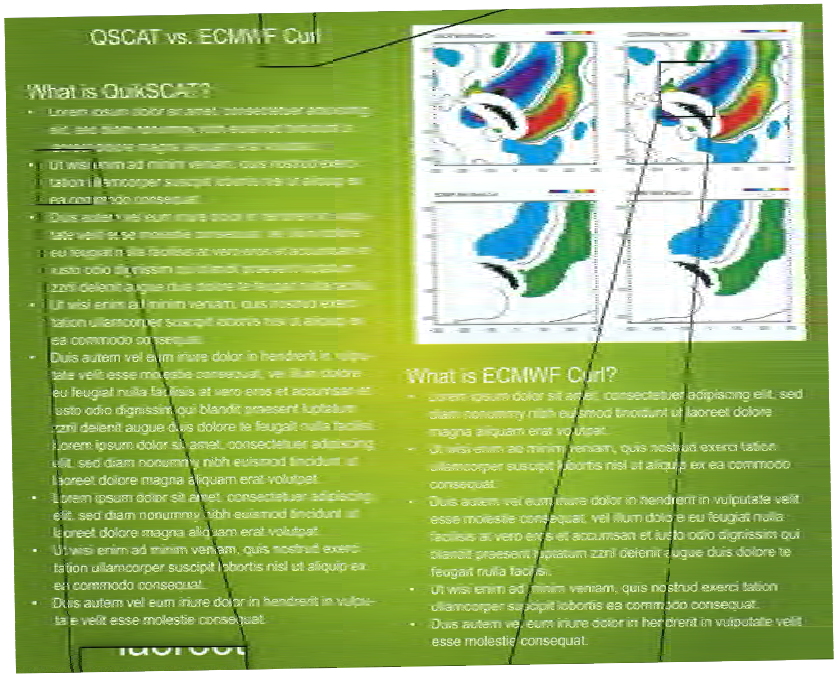
Example 3 – Big Data

add some discussion

Data distribution, potential for over fitting and smoothing

Of course, the figure is hard to read, even in the printed publication

Look at the website referenced, did the research attempt to make a video of changes with time?



Example set 4

from The Oceanographic Society (2005) Scientifically Speaking: Tips for Preparing and Delivering Scientific Talks and Using Visual Aids. 24p.

modified from page 14 Example of a “Bad” PowerPoint slide

bad slide for many reasons – a poorly designed and executed visual aid suggests a lack of professionalism, preparation, and commitment to the audience and science. Poor visuals degrade communication.

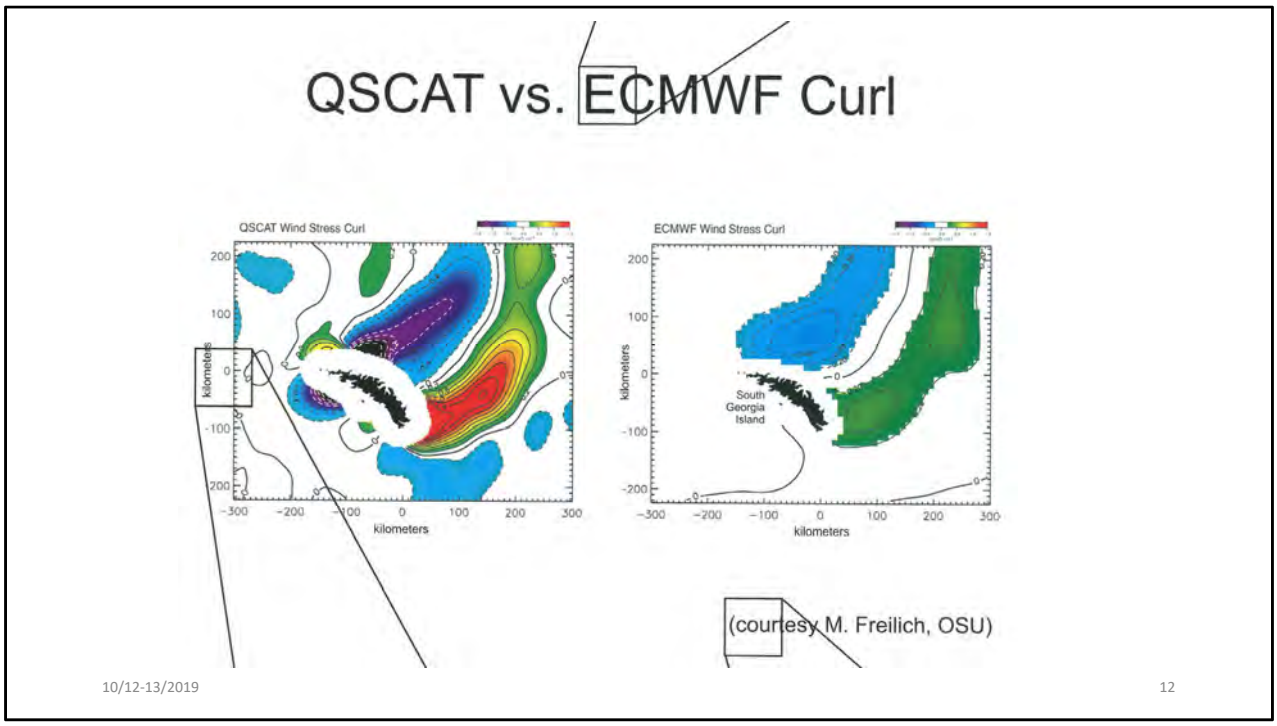
Title too small

green background is distracting and decreases readability

There is too much text – this is what the presenter should be developing and discussing, it is not visual aid

The graphics are too small, labels are unreadable

The source of the graphics should be referenced



Oceanographic Society publication, modified from page 15 Example of a "Good" PowerPoint slide

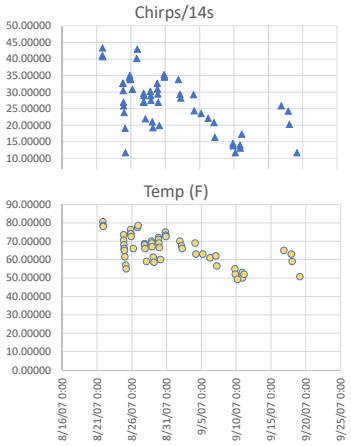
Example of a much better slide

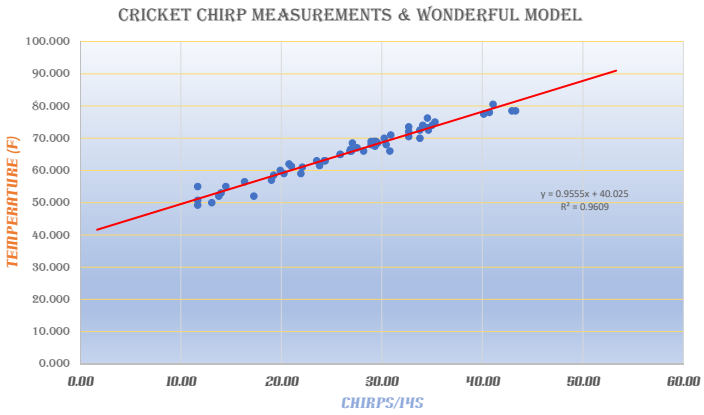
Because they automatically assume center stage, it is vitally important that all visual aids clarify and support your talk in an attractive, comprehensible manner or they will detract from your talk and loose the audience attention

The slide title is 44pt sans serif font, which makes it very legible
The white background makes the text and graphics much more legible in all rooms
Only two graphics are used, which focuses the presentation on one key point (the pseudo "QSCAT vs. ECMWF Curl")
The axes are labeled and almost readable – in fact, the minor tick marks are unnecessary and the major axes should still be simplified.
The color bar legend should only be shown once
The source of the data is acknowledged

Figure interactive development

Can we improve this data visualization?





10/12-13/2019

13

Using Excel graphics capabilities we will revise this slide during the session

crickets chirp data from last year (see notes below) – the slide starts out pretty ugly, though PowerPoint provides lots of hints and suggestions that could make it much worse, in Dave's opinion (and others)

???? *add some simple statistical modeling code results? -- DRJ will work on this early this week*

Example from "The Old Farmer's Almanac": if you live near crickets, you can estimate the nighttime outdoor temperature in degrees Fahrenheit by counting the number of cricket chirps in 14 seconds and adding 40 to that number.

First examine data collection and relationships – crickets

Crickets and temperature are a great example of correlated data

- What does correlated imply?
- How can we examine and test correlations using mathematics?

Can computer simulations generate data for such analysis?

The observations were recorded by Dr. Margaret LeMone in Boulder, Colorado, over a 30 day period in August and September, 2007 [2].

The measurements were originally in chirps per 30 seconds; the column for chirps per 14 seconds was derived from the original data.

What further scientific questions can you ask? Are those questions testable?

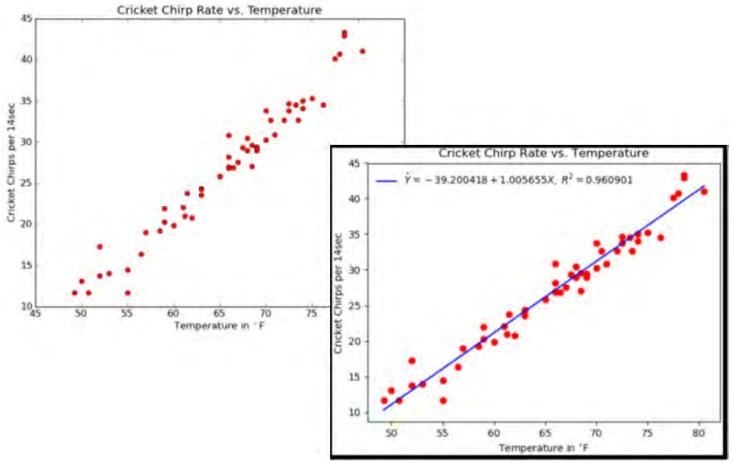
Dave's approach to improving the slide:

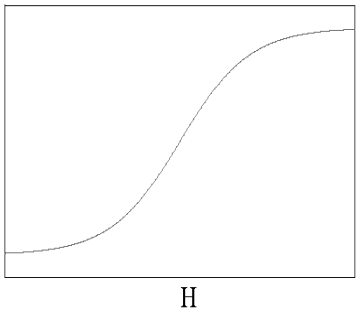
Fix the fonts, significant figures, colors & remove graph background fill, font sizes, marker styles

Do we need the two left figures, or should they be on an earlier background slide

Add acknowledgement to Dr. Margaret LeMone for data

Did we do better data visualization?





10/12-13/2019

14

Using Excel graphics capabilities
compare to previous exercise

The observations were recorded by Dr. Margaret LeMone in Boulder, Colorado, over a 30 day period in August and September, 2007 [2].
The measurements were originally in chirps per 30 seconds; the column for chirps per 14 seconds was derived from the original data.
What further scientific questions can you ask? Are those questions testable?

Summary

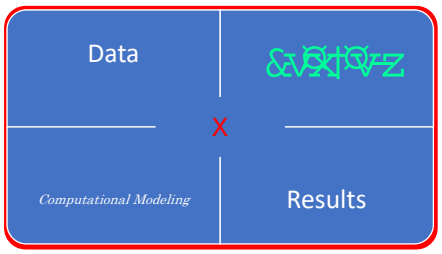
to help making a good visualization figure

3”x5” sized figure at arms length allows you to check readability

or printing 4 slides per landscape page in PowerPoint is equivalent

Fonts need to be simple

Use consistent font throughout the report, poster, and slide set



Add other directions from Championing Science
And visualization books by Tufte

Blue vs red perception
Reading left to right, however some aspects draw attention (like the red X in the slide graphic)

Visualization is more than one figure for a Supercomputing Challenge Project

The project definition, including major question, sub-questions, & team

Background information, including resources & mentors

Computer modeling approach(es) and development
this is a computing challenge

Modeling results, approach refinement, testing & evaluation

Conclusions

Acknowledgements

References

Code Documentation

"A good scientist doesn't stop once {*they}* get the answer to a question, {*they}* look at it repeated from different angles." from Fisher (2012)

Now take a broader view of visualization for the Supercomputing Challenge project:
This covers the project itself, the written report, a poster, and an oral presentation

From: Fisher, Ken (2012) The only three questions that still count: Investing by knowing what others don't. John Wiley & Sons, Inc. 348pp.

Whenever you're confirming …, try it from a fresh angle. Go crazy. Be creative. Flip things on their heads, backward and inside out. Hack them up and go over their guts. Instead of trying to be intuitive, think counter intuitively – which may turn out to be much more intuitive

As you strike out on your own, testing your own mythology, make sure you're thorough. A good scientist doesn't stop once he gets the answer to a question; he looks at it repeatedly from different angles.

Reference to report structure suggestions and examples on Supercomputing Challenge website

Poster format

for Supercomputing Challenge project

What are the key components?

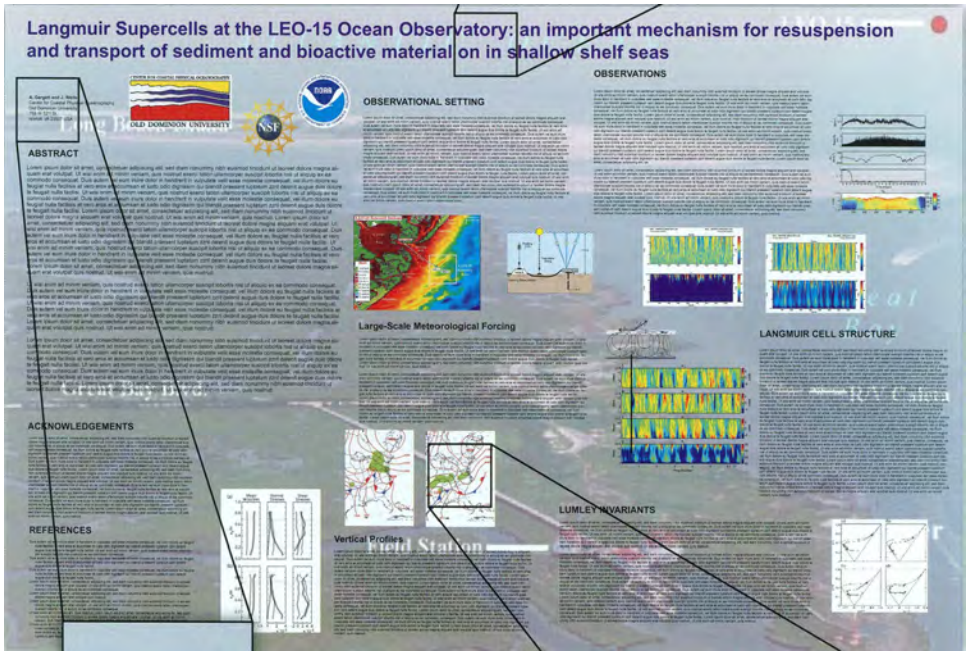
Dave's initial structure suggestion

Drawing to be inserted here

Following from the introduction discussion (and identification of whether the students have previously made and presented a poster):
Identify and discuss key components of a poster with students in the session
then what are levels of emphasis (in general percentages of poster space coverage)

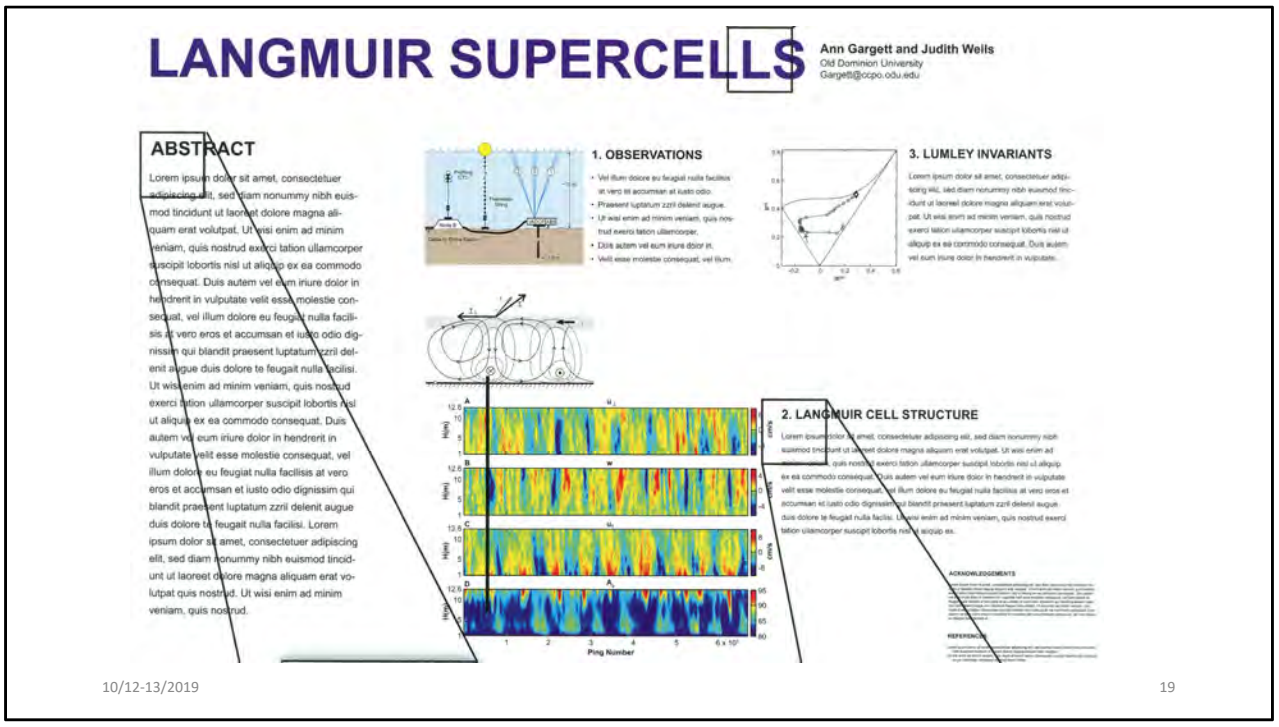
Draw out suggested structure

quarters, emphasizing computation approach, code development, and computing results in at least half, if not three quarters of the poster



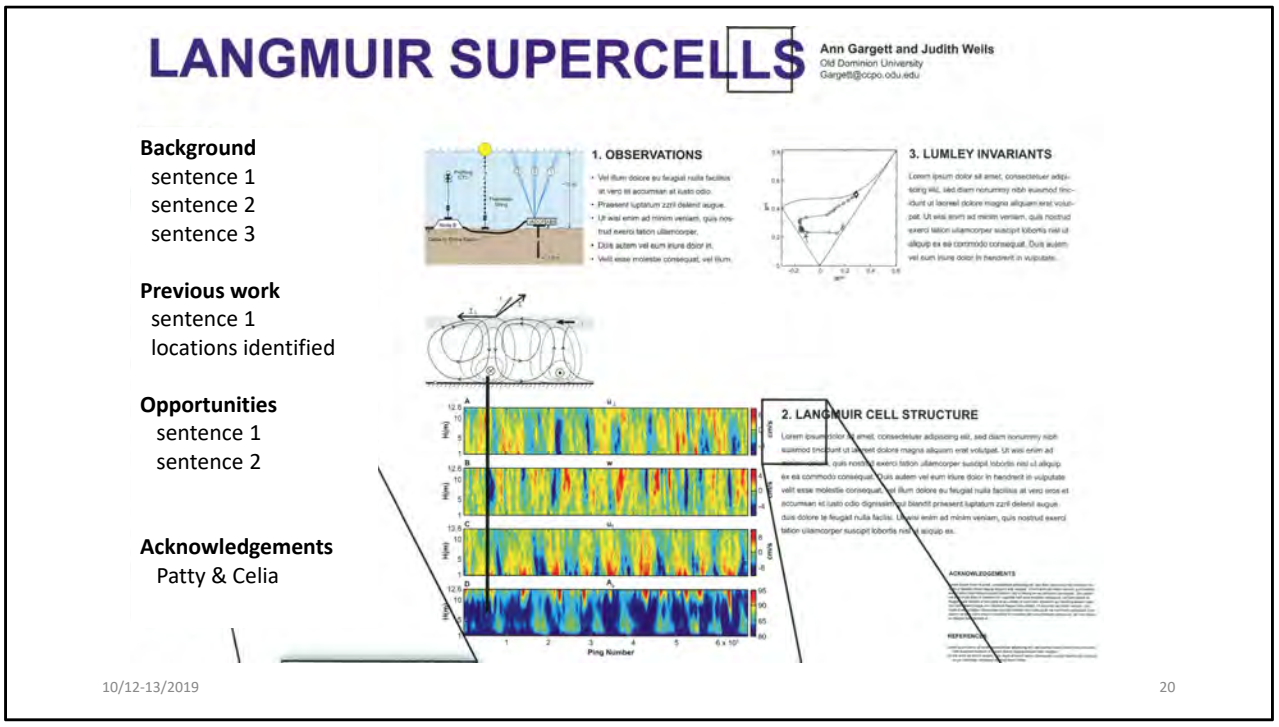
Oceanographic Society publication, page 2 Example of a “Bad” Poster – poor

Add notes



Oceanographic Society publication, page 3 Example of a “Good” Poster – good

Add notes from publication



Oceanographic Society publication, modified from page 3 Example of a “Good” Poster – better

Add notes from publication discussion

Remove abstract and replace with summary of problems and approach

For Supercomputing Challenge, your written project report will be available, and you should have it at the poster session (EXPO) to refer to for more detailed questions and answers

Supercomputing Challenge

Summary thoughts – A suite of communication approaches

Steve Ballmer doesn't want to see slide decks: too inefficient

Another example of how it may be more effective to distribute handouts/reading material before a presentation versus enslaving a meeting to a deck of slides comes from Steve Ballmer From "Meetings, Version 2.0, at Microsoft", in the New York Times' Corner Office column (http://www.nytimes.com/2009/05/17/business/17corner.html):

New York Times: "What's it like to be in a meeting run by Steve Ballmer?"

Steve Ballmer: "I've changed that, really in the last couple years. The mode of Microsoft meetings used to be: You come with something we haven't seen in a slide deck or presentation. You deliver the presentation. You probably take what I will call "the long and winding road." You take the listener through your path of discovery and exploration, and you arrive at a conclusion.

That's kind of the way I used to like to do it, and the way Bill [Gates] used to kind of like to do it. And it seemed like the best way to do it, because if you went to the conclusion first, you'd get: "What about this? Have you thought about this?" So people naturally tried to tell you all the things that supported the decision, and then tell you the decision.

I decided that's not what I want to do anymore. I don't think it's productive. I don't think it's efficient. I get impatient. So most meetings nowadays, you send me the materials and I read them in advance. And I can come in and say: "I've got the following four questions. Please don't present the deck." That lets us go, whether they've organized it that way or not, to the recommendation. And if I have questions about the long and winding road and the data and the supporting evidence, I can ask them. But it gives us greater focus."

Captured from https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0002PP

A final recommendation about the communication approaches that will be used during a 2019-2020 Supercomputing Challenge project

When project presentations are made in April, you will have completed a written report. So the poster and oral presentations are visual aids for your talking about the project.

Yes, an oral presentation will not be an open discussion like Steve Ballmer identifies (I hope you will have plenty of those opportunities in the future!). So use your poster and visual aids for an oral presentation as supporting information for what you are talking about, as discussed in the previous slides.

References

Visualization references:

The Oceanographic Society (2005) Scientifically Speaking: Tips for Preparing and Delivering Scientific Talks and Using Visual Aids. 24p. https://tos.org/scientifically-speaking

Aines, R.D. & Aines, A.L. (2019) Championing Science: Communicating Your Ideas to Decision Makers. University of California Press 249p.

Mermin, N.D. (1992) What's Wrong with Those Talks? Physics Today 45(11) 9-11 [https://doi.org/10.1063/1.2809861]

Fisher, Ken (2012) The only three questions that still count: Investing by knowing what others don't. John Wiley & Sons, Inc. 348pp.

Supercomputing Challenge website -- https://supercomputingchallenge.org/19-20/index.php

Wikipedia: Data Visualization – https://en.wikipedia.org/wiki/Data_visualization

Wikipedia: Scientific Visualization -- https://en.wikipedia.org/wiki/Scientific_visualization

Data visualization beginner's guide -- https://www.tableau.com/learn/articles/data-visualization

Scientific paper references:

Mike Tercek (2019) Nowcasting & Forecasting Fire Severity in Yellowstone, Yellowstone Science 27(1) 27-33

Andrew M. Ray, David P. Thoma, Kristin L. Legg, David M. Diamond, & Andrew J. Hasen (2019) Assessing the Ecological Health of the Greater Yellowstone Ecosystem. Yellowstone Science 27(1) 18-21

A.B. Toth *et. al.* (2019) Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. Science 365(6459) 1305-1308

Roberto Spina (2019) Big data and artificial intelligence analytics in Geosciences: Promises and Potential. GSA Today 29(1) 42-43

ADD to these notes

Specific links on Supercomputing Challenge website, including previous reports
Tufte books & pamphlet

The Yellowstone Science articles by Tercek and Ray et.al. are available at
https://www.nps.gov/articles/series.htm?id=8A1106D1-CBC3-15BF-85C5DD6B6D66D4EF