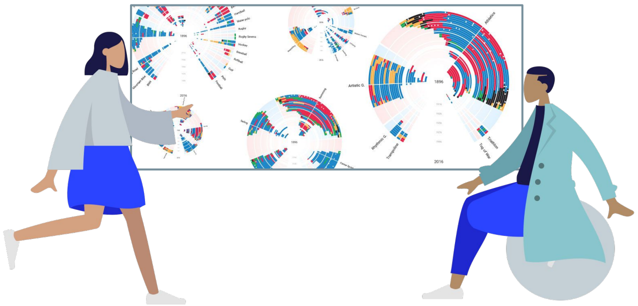
artur.trindade
beatriz.malveiro
david.polido
liliana.fernandes
rita.costa
@feedzai.com

feedzai

Data Visualization **Show & Tell**



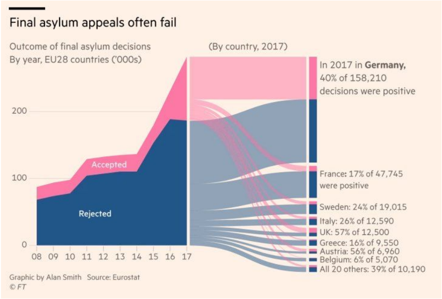
data visualization
lisboa **#vislis**

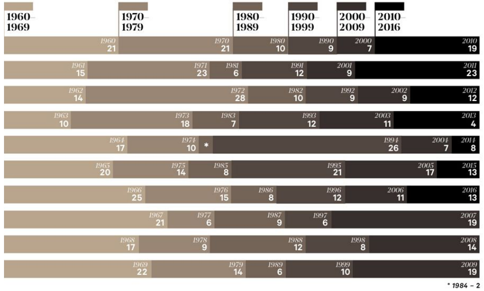
Agenda

- Welcome
- Critique
- Visualizing Uncertainty
- Community Chart Clinic
- End of Event


Critique

Critique


Final asylum appeals often fail
Outcome of final asylum decisions
By year, EU28 countries ('000s)
(By country, 2017)
200
100
0
Accepted
Rejected
08 09 10 11 12 13 14 15 16 17
In 2017 in Germany, 40% of 158,210 decisions were positive
France: 17% of 47,745 were positive
Sweden: 24% of 19,015
Italy: 26% of 12,590
UK: 57% of 12,500
Greece: 16% of 9,550
Austria: 56% of 6,960
Belgium: 6% of 5,070
All 20 others: 39% of 10,190
Graphic by Alan Smith Source: Eurostat
© FT

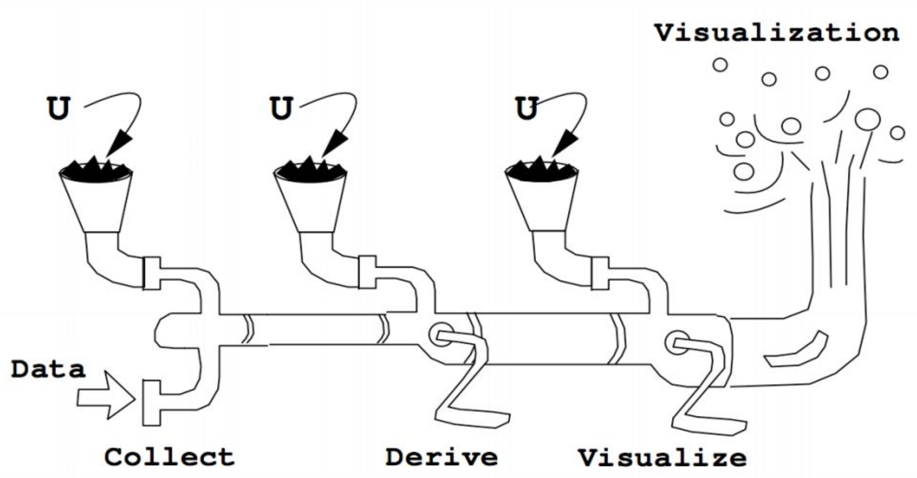

1960 1969
1970 1979
1980 1989
1990 1999
2000 2009
2010 2016
* 1984 – 2

Visualizing Uncertainty

Uncertainty What is it?

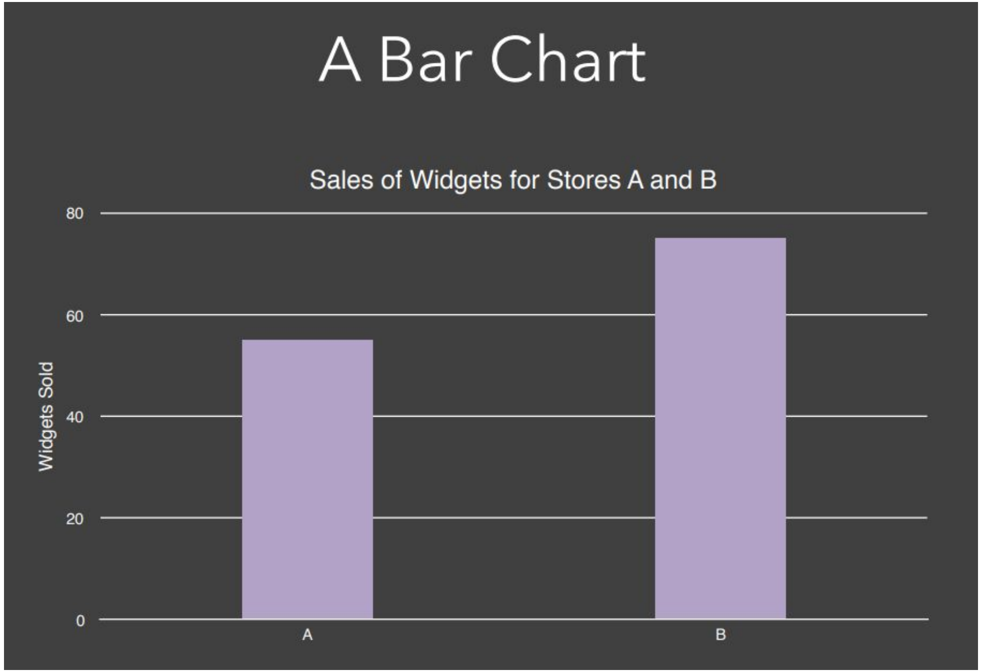
What “Uncertainty” Can Mean

- Doubt
- Risk
- Variability
- Error
- Lack of Knowledge
- Hedging



Pang et al. Approaches to Uncertainty Visualization. The Visual Computer, 1997.

https://courses.cs.washington.edu/courses/cse512/19sp/lectures/CSE512-Uncertainty.pdf



Uncertainty Sources

- Measurement
- Forecast
- Model
- Decision

https://courses.cs.washington.edu/courses/cse512/19sp/lectures/CSE512-Uncertainty.pdf

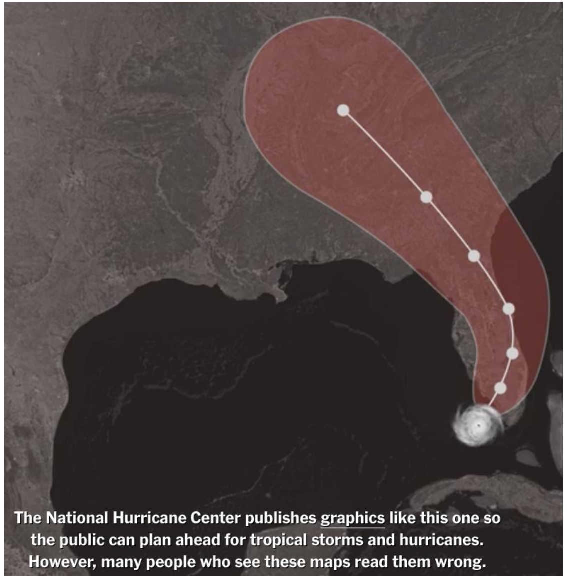
Uncertainty Sources

- Measurement *"We're not sure what the data are"*
- Forecast *"We're not sure what will happen to the data next"*
- Model *"We're not sure how the data fit together"*
- Decision *"We're not sure what to do with the data"*

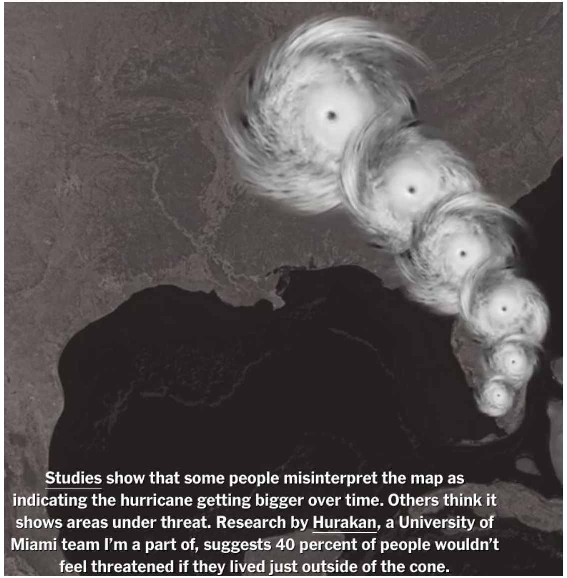
https://courses.cs.washington.edu/courses/cse512/19sp/lectures/CSE512-Uncertainty.pdf

The New York Times

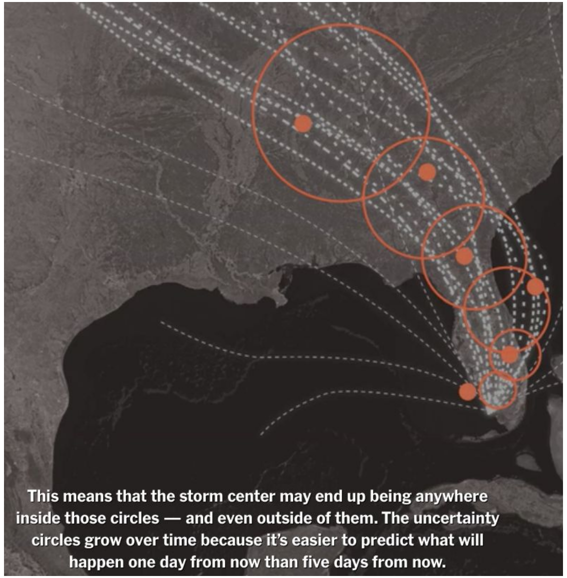
How charts look like



How people read them



How it should be read

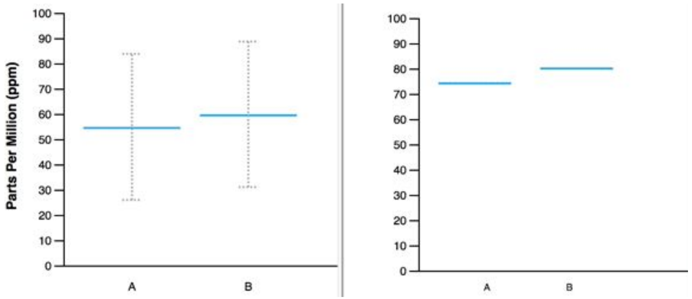


https://www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html

Visualizing Uncertainty | Animation

The New York Times [Live US Presidential Forecast (2016)](#)

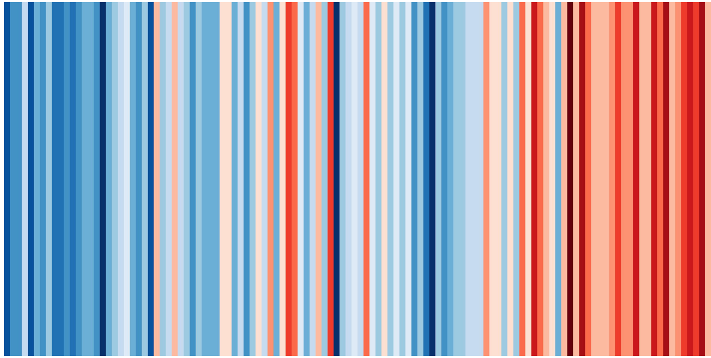
Midwest Uncertainty Collective (MU) [HOPS plots](#)



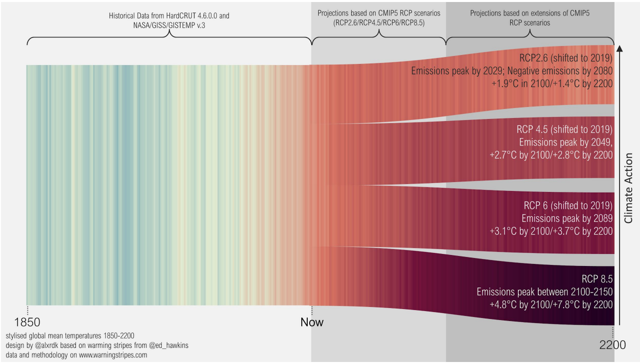
Uncertainty Visualizing Uncertainty | Multiple outcomes

#ShowYourStripes

Warming Stripes for Portugal from 1901-2018



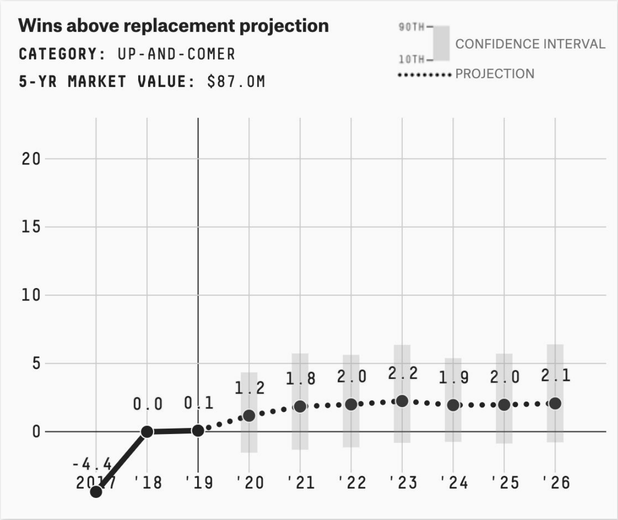
Warming stripes with projections



Uncertainty **Visualizing Uncertainty | Taxonomy**

FLOWINGDATA [Visualizing the Uncertainty in Data]

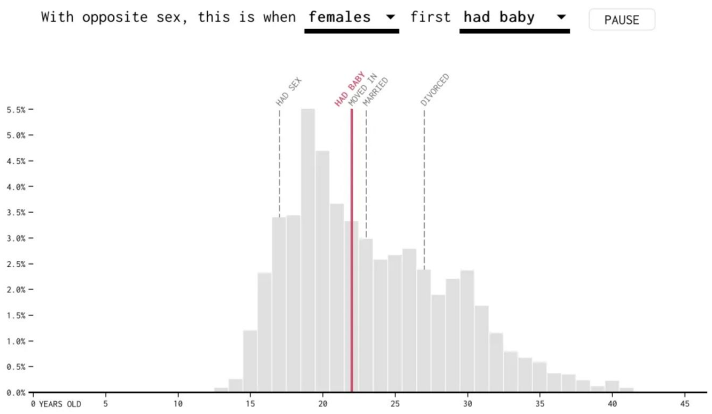
Ranges | Distributions | Multiple Outcomes | Simulations | Obscurity | Words



With **ranges** we see that a mean or median represents only part of an estimate but lot of people don't understand the concept of confidence intervals or what standard error bars are, so you need to explain clearly with annotation.

FLOWINGDATA [Visualizing the Uncertainty in Data]

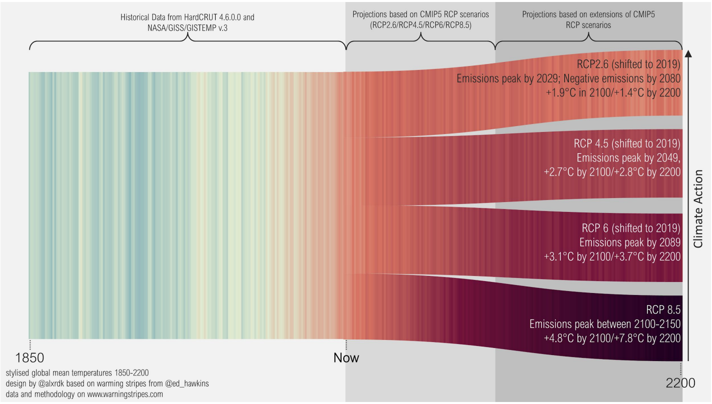
Ranges Distributions Multiple Outcomes Simulations Obscurity Words



By showing the variation in a sample, the reader can make a more educated judgement about whether a sample is trustworthy. But again, many people don't understand **distributions**, so you need to explain what's going on.

FLOWINGDATA [Visualizing the Uncertainty in Data]

Ranges Distributions **Multiple Outcomes** Simulations Obscurity Words

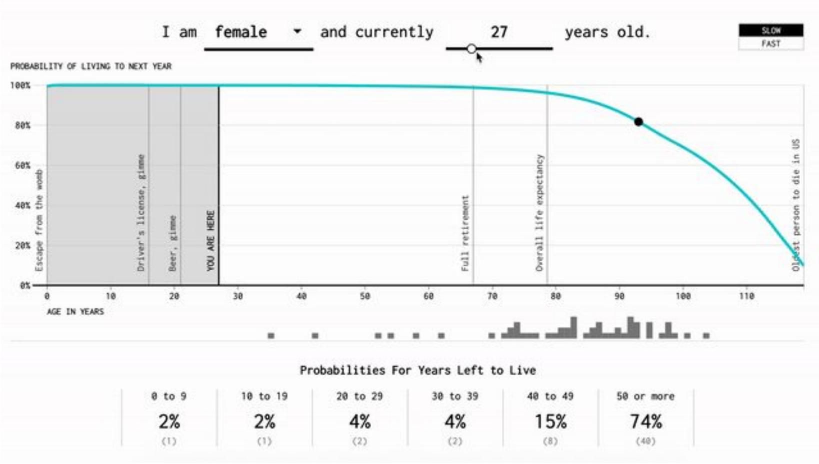


Uncertainty is displayed more explicitly. People can see that there is no set path, and instead they see a bunch of possible paths.

Uncertainty Visualizing Uncertainty | Taxonomy

FLOWINGDATA [Visualizing the Uncertainty in Data]

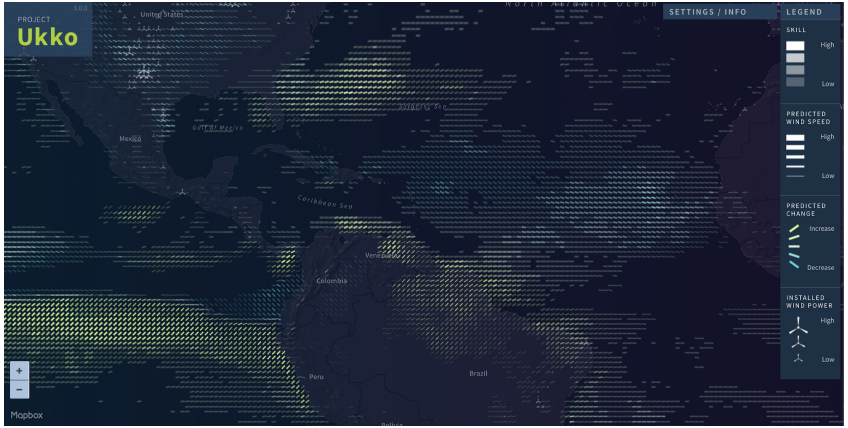
Ranges Distributions Multiple Outcomes **Simulations** Obscurity Words



When data appears all at once or in aggregate, it can be a challenge for many to interpret results and link it back to what the data actually represents. By showing **simulations**, you get a sense of build-up and a link with individual outcomes.

FLOWINGDATA [Visualizing the Uncertainty in Data]

Ranges Distributions Multiple Outcomes Simulations **Obscurity** Words



Lines represent wind predictions, and opacity represents the strength of the predictions.

If you're less certain about an estimate, make it less visually prominent. The data that's less up in the air gets more attention as a result.

Ranges Distributions Multiple Outcomes Simulations Obscurity Words

Maybe visualization isn't what you're looking for at all. After all, you don't have to visualize everything! You can add uncertainty to your writing by **avoiding absolutes when you describe numbers**. Treat estimates as such when you use them, and account for the uncertainty in the numbers.

Midwest Uncertainty Collective (MU)

Jessica Hullman
@JessicaHullman

Matthew Kay
@mjskay

Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

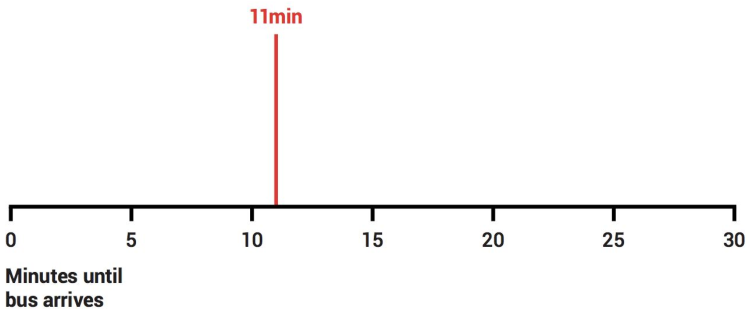
UW Interactive Data Lab

Do I have time to get a coffee? ☕

Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

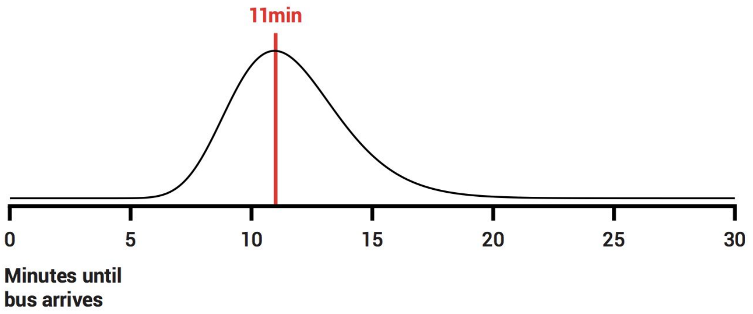
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

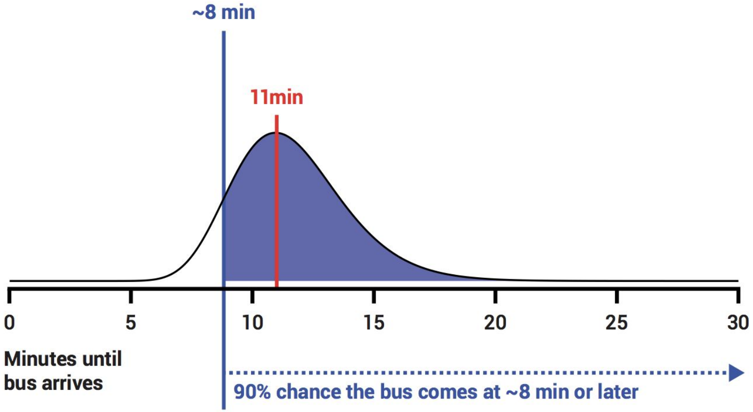
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

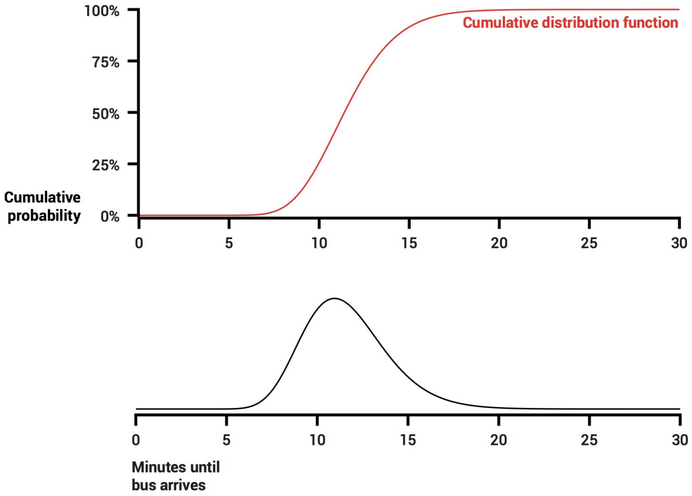
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

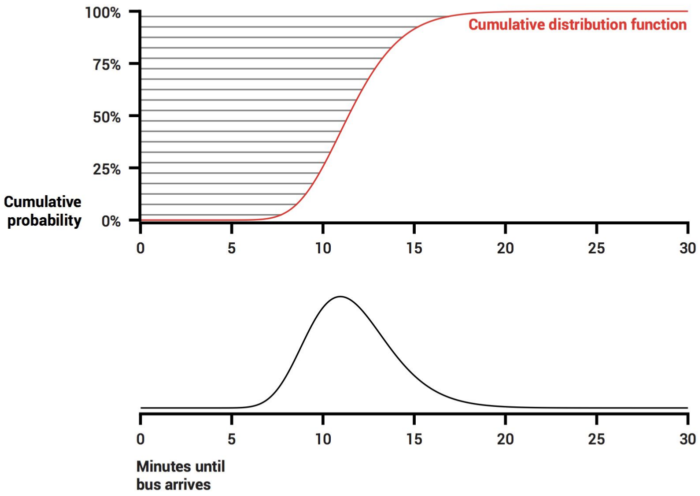
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

UW Interactive Data Lab

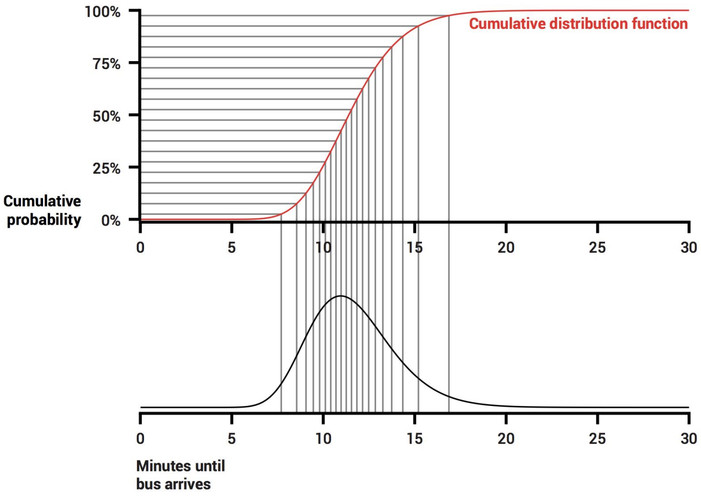


feedzai

Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

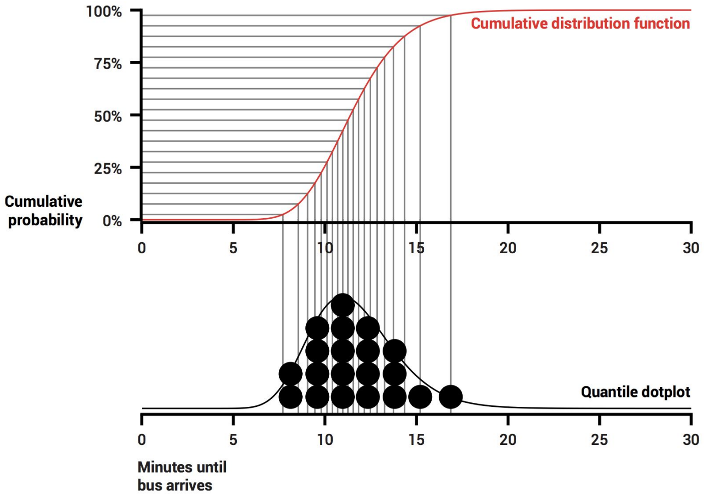
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

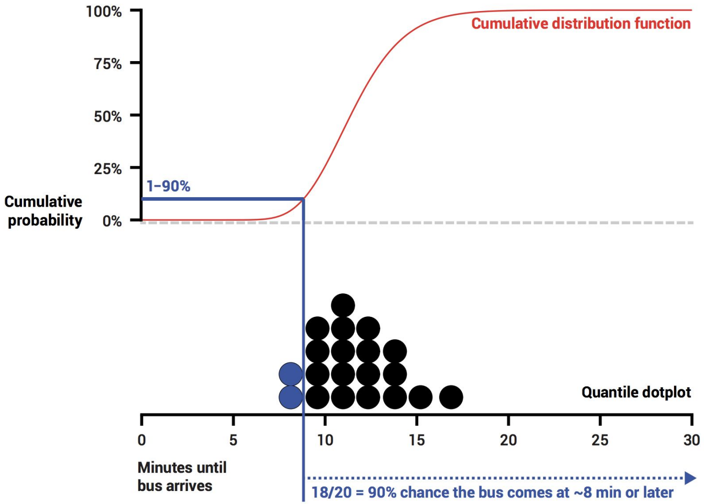
UW Interactive Data Lab



Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

UW Interactive Data Lab

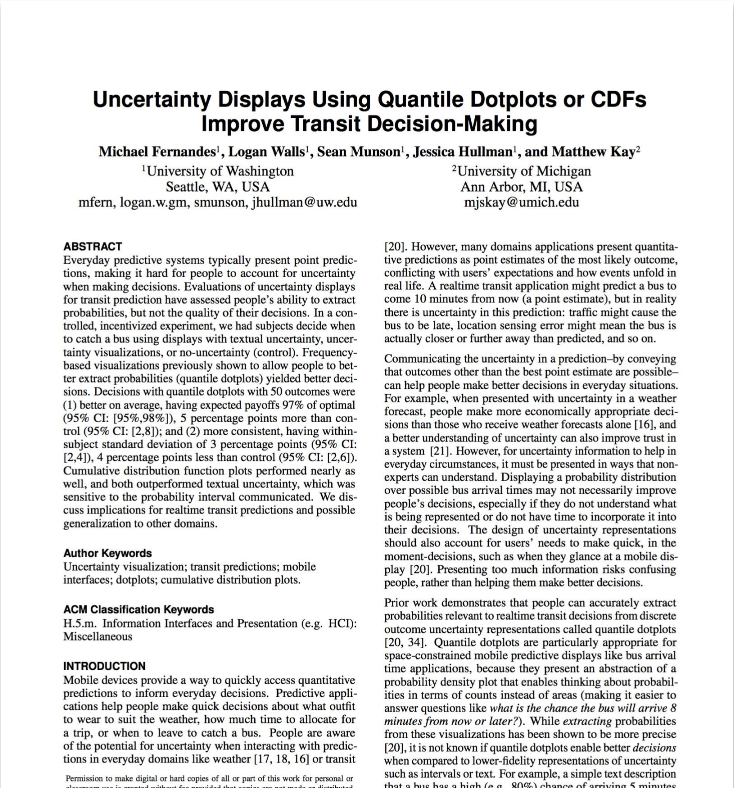


feedzai

Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

UW Interactive Data Lab



Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making

Michael Fernandes[1], Logan Walls[1], Sean Munson[1], Jessica Hullman[1], and Matthew Kay[2]

[1]University of Washington
Seattle, WA, USA
mfern, logan.w.gm, smunson, jhullman@uw.edu

[2]University of Michigan
Ann Arbor, MI, USA
mjskay@umich.edu

ABSTRACT

Everyday predictive systems typically present point predictions, making it hard for people to account for uncertainty when making decisions. Evaluations of uncertainty displays for transit prediction have assessed people's ability to extract probabilities, but not the quality of their decisions. In a controlled, incentivized experiment, we had subjects decide when to catch a bus using displays with textual uncertainty, uncertainty visualizations, or no-uncertainty (control). Frequency-based visualizations previously shown to allow people to better extract probabilities (quantile dotplots) yielded better decisions. Decisions with quantile dotplots with 50 outcomes were (1) better on average, having expected payoffs 97% of optimal (95% CI: [95%,98%]), 5 percentage points more than control (95% CI: [2,8]); and (2) more consistent, having within-subject standard deviation of 3 percentage points (95% CI: [2,4]), 4 percentage points less than control (95% CI: [2,6]). Cumulative distribution function plots performed nearly as well, and both outperformed textual uncertainty, which was sensitive to the probability interval communicated. We discuss implications for realtime transit predictions and possible generalization to other domains.

Author Keywords

Uncertainty visualization; transit predictions; mobile interfaces; dotplots; cumulative distribution plots.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

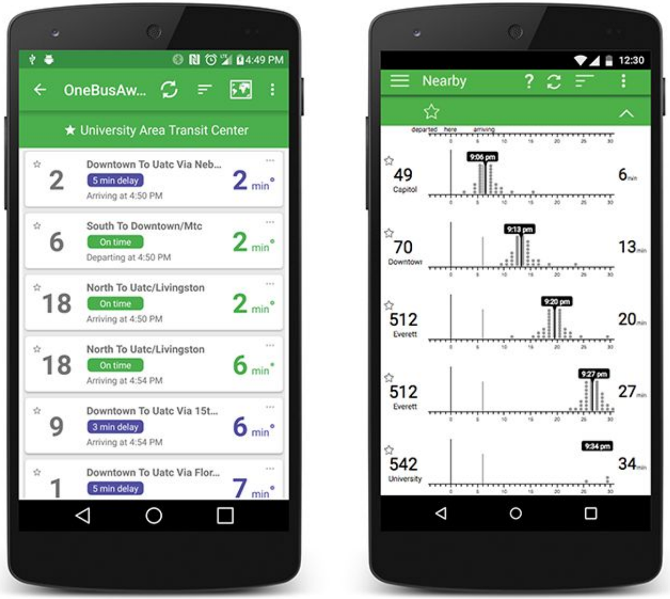
INTRODUCTION

Mobile devices provide a way to quickly access quantitative predictions to inform everyday decisions. Predictive applications help people make quick decisions about what outfit to wear to suit the weather, how much time to allocate for a trip, or when to leave to catch a bus. People are aware of the potential for uncertainty when interacting with predictions in everyday domains like weather [17, 18, 16] or transit [20]. However, many domains applications present quantitative predictions as point estimates of the most likely outcome, conflicting with users' expectations and how events unfold in real life. A realtime transit application might predict a bus to come 10 minutes from now (a point estimate), but in reality there is uncertainty in this prediction: traffic might cause the bus to be late, location sensing error might mean the bus is actually closer or further away than predicted, and so on.

Communicating the uncertainty in a prediction–by conveying that outcomes other than the best point estimate are possible–can help people make better decisions in everyday situations. For example, when presented with uncertainty in a weather forecast, people make more economically appropriate decisions than those who receive weather forecasts alone [16], and a better understanding of uncertainty can also improve trust in a system [21]. However, for uncertainty information to help in everyday circumstances, it must be presented in ways that non-experts can understand. Displaying a probability distribution over possible bus arrival times may not necessarily improve people's decisions, especially if they do not understand what is being represented or do not have time to incorporate it into their decisions. The design of uncertainty representations should also account for users' needs to make quick, in the moment-decisions, such as when they glance at a mobile display [20]. Presenting too much information risks confusing people, rather than helping them make better decisions.

Prior work demonstrates that people can accurately extract probabilities relevant to realtime transit decisions from discrete outcome uncertainty representations called quantile dotplots [20, 34]. Quantile dotplots are particularly appropriate for space-constrained mobile predictive displays like bus arrival time applications, because they present an abstraction of a probability density plot that enables thinking about probabilities in terms of counts instead of areas (making it easier to answer questions like *what is the chance the bus will arrive 8 minutes from now or later?*). While *extracting* probabilities from these visualizations has been shown to be more precise [20], it is not known if quantile dotplots enable better *decisions* when compared to lower-fidelity representations of uncertainty such as intervals or text. For example, a simple text description

Permission to make digital or hard copies of all or part of this work for personal or

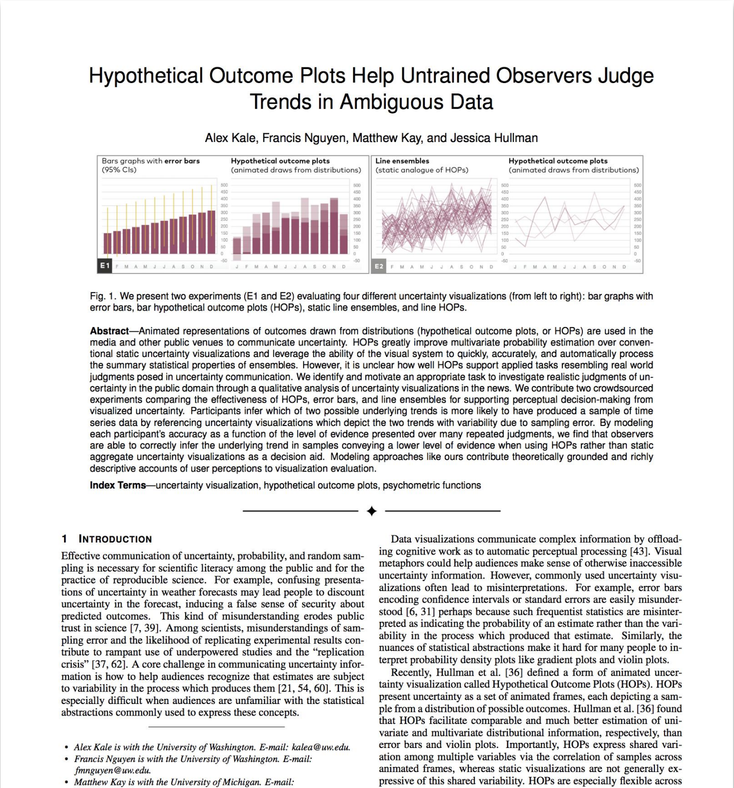


data visualization
lisboa #vislis

Uncertainty State of the Art

Midwest Uncertainty Collective (MU)

UW Interactive Data Lab



Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data

Alex Kale, Francis Nguyen, Matthew Kay, and Jessica Hullman

Bars graphs with error bars (95% CIs) | Hypothetical outcome plots (animated draws from distributions) | Line ensembles (static analogue of HOPs) | Hypothetical outcome plots (animated draws from distributions)

Fig. 1. We present two experiments (E1 and E2) evaluating four different uncertainty visualizations (from left to right): bar graphs with error bars, bar hypothetical outcome plots (HOPs), static line ensembles, and line HOPs.

Abstract—Animated representations of outcomes drawn from distributions (hypothetical outcome plots, or HOPs) are used in the media and other public venues to communicate uncertainty. HOPs greatly improve multivariate probability estimation over conventional static uncertainty visualizations and leverage the ability of the visual system to quickly, accurately, and automatically process the summary statistical properties of ensembles. However, it is unclear how well HOPs support applied tasks resembling real world judgments posed in uncertainty communication. We identify and motivate an appropriate task to investigate realistic judgments of uncertainty in the public domain through a qualitative analysis of uncertainty visualizations in the news. We contribute two crowdsourced experiments comparing the effectiveness of HOPs, error bars, and line ensembles for supporting perceptual decision-making from visualized uncertainty. Participants infer which of two possible underlying trends is more likely to have produced a sample of time series data by referencing uncertainty visualizations which depict the two trends with variability due to sampling error. By modeling each participant's accuracy as a function of the level of evidence presented over many repeated judgments, we find that observers are able to correctly infer the underlying trend in samples conveying a lower level of evidence when using HOPs rather than static aggregate uncertainty visualizations as a decision aid. Modeling approaches like ours contribute theoretically grounded and richly descriptive accounts of user perceptions to visualization evaluation.

Index Terms—uncertainty visualization, hypothetical outcome plots, psychometric functions

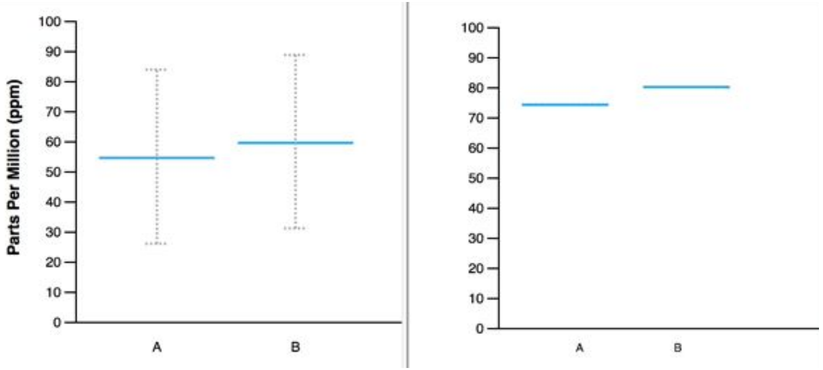
1 INTRODUCTION

Effective communication of uncertainty, probability, and random sampling is necessary for scientific literacy among the public and for the practice of reproducible science. For example, confusing presentations of uncertainty in weather forecasts may lead people to discount uncertainty in the forecast, inducing a false sense of security about predicted outcomes. This kind of misunderstanding erodes public trust in science [7, 39]. Among scientists, misunderstandings of sampling error and the likelihood of replicating experimental results contribute to rampant use of underpowered studies and the "replication crisis" [37, 62]. A core challenge in communicating uncertainty information is how to help audiences recognize that estimates are subject to variability in the process which produces them [21, 54, 60]. This is especially difficult when audiences are unfamiliar with the statistical abstractions commonly used to express these concepts.

Data visualizations communicate complex information by offloading cognitive work as to automatic perceptual processing [43]. Visual metaphors could help audiences make sense of otherwise inaccessible uncertainty information. However, commonly used uncertainty visualizations often lead to misinterpretations. For example, error bars encoding confidence intervals or standard errors are easily misunderstood [6, 31] perhaps because such frequentist statistics are misinterpreted as indicating the probability of an estimate rather than the variability in the process which produced that estimate. Similarly, the nuances of statistical abstractions make it hard for many people to interpret probability density plots like gradient plots and violin plots.

Recently, Hullman et al. [36] defined a form of animated uncertainty visualization called Hypothetical Outcome Plots (HOPs). HOPs present uncertainty as a set of animated frames, each depicting a sample from a distribution of possible outcomes. Hullman et al. [36] found that HOPs facilitate comparable and much better estimation of univariate and multivariate distributional information, respectively, than error bars and violin plots. Importantly, HOPs express shared variation among multiple variables via the correlation of samples across animated frames, whereas static visualizations are not generally expressive of this shared variability. HOPs are especially flexible across

• Alex Kale is with the University of Washington. E-mail: kalea@uw.edu.
• Francis Nguyen is with the University of Washington. E-mail: fmnguyen@uw.edu.
• Matthew Kay is with the University of Michigan. E-mail:



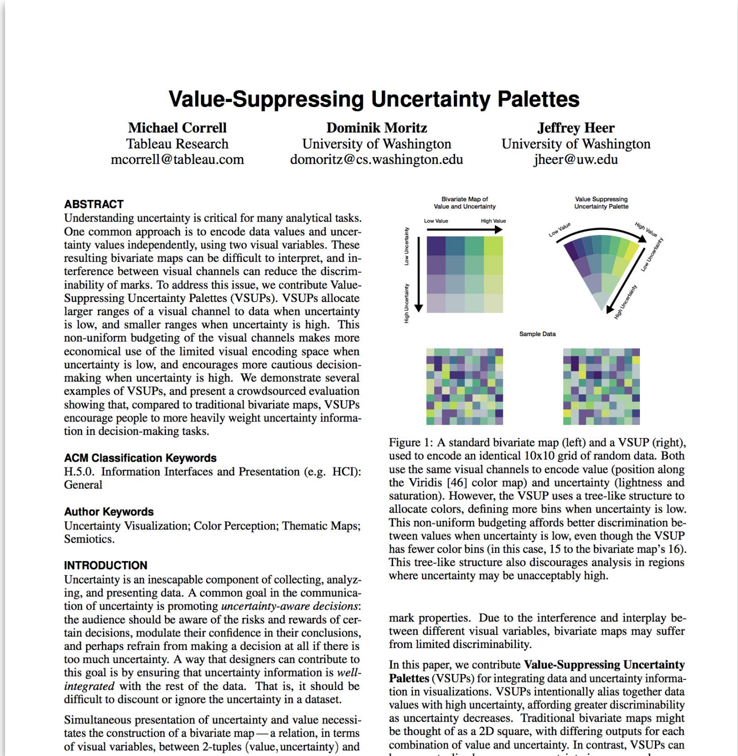
"Experience uncertainty"

feedzai

Uncertainty **State of the Art**

UW Interactive Data Lab

TABLEAU RESEARCH



Value-Suppressing Uncertainty Palettes

Michael Correll
Tableau Research
mcorrell@tableau.com

Dominik Moritz
University of Washington
domoritz@cs.washington.edu

Jeffrey Heer
University of Washington
jheer@uw.edu

ABSTRACT
Understanding uncertainty is critical for many analytical tasks. One common approach is to encode data values and uncertainty values independently, using two visual variables. These resulting bivariate maps can be difficult to interpret, and interference between visual channels can reduce the discriminability of marks. To address this issue, we contribute Value-Suppressing Uncertainty Palettes (VSUPs). VSUPs allocate larger ranges of a visual channel to data when uncertainty is low, and smaller ranges when uncertainty is high. This non-uniform budgeting of the visual channels makes more economical use of the limited visual encoding space when uncertainty is low, and encourages more cautious decision-making when uncertainty is high. We demonstrate several examples of VSUPs, and present a crowdsourced evaluation showing that, compared to traditional bivariate maps, VSUPs encourage people to more heavily weight uncertainty information in decision-making tasks.

ACM Classification Keywords
H.5.0. Information Interfaces and Presentation (e.g. HCI): General

Author Keywords
Uncertainty Visualization; Color Perception; Thematic Maps; Semiotics.

INTRODUCTION
Uncertainty is an inescapable component of collecting, analyzing, and presenting data. A common goal in the communication of uncertainty is promoting *uncertainty-aware decisions*: the audience should be aware of the risks and rewards of certain decisions, modulate their confidence in their conclusions, and perhaps refrain from making a decision at all if there is too much uncertainty. A way that designers can contribute to this goal is by ensuring that uncertainty information is *well-integrated* with the rest of the data. That is, it should be difficult to discount or ignore the uncertainty in a dataset.

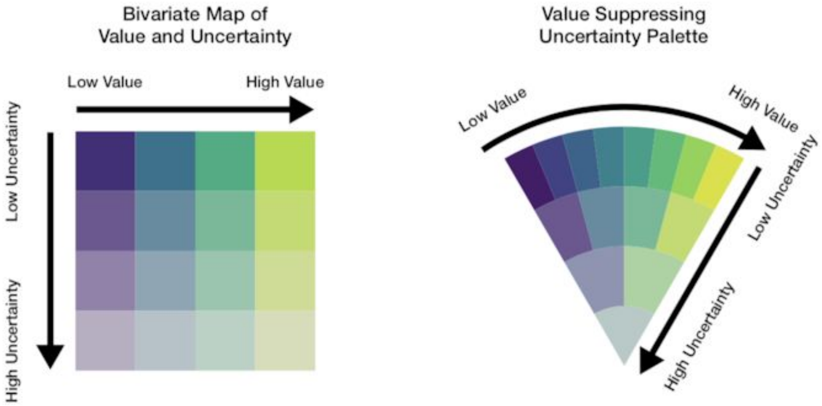
Simultaneous presentation of uncertainty and value necessitates the construction of a bivariate map — a relation, in terms of visual variables, between 2-tuples (value, uncertainty) and

Bivariate Map of Value and Uncertainty | Value Suppressing Uncertainty Palette | Sample Data

Figure 1: A standard bivariate map (left) and a VSUP (right), used to encode an identical 10x10 grid of random data. Both use the same visual channels to encode value (position along the Viridis [46] color map) and uncertainty (lightness and saturation). However, the VSUP uses a tree-like structure to allocate colors, defining more bins when uncertainty is low. This non-uniform budgeting affords better discrimination between values when uncertainty is low, even though the VSUP has fewer color bins (in this case, 15 to the bivariate map's 16). This tree-like structure also discourages analysis in regions where uncertainty may be unacceptably high.

mark properties. Due to the interference and interplay between different visual variables, bivariate maps may suffer from limited discriminability.

In this paper, we contribute **Value-Suppressing Uncertainty Palettes** (VSUPs) for integrating data and uncertainty information in visualizations. VSUPs intentionally alias together data values with high uncertainty, affording greater discriminability as uncertainty decreases. Traditional bivariate maps might be thought of as a 2D square, with differing outputs for each combination of value and uncertainty. In contrast, VSUPs can



Uncertainty State of the Art

UW Interactive Data Lab

Previously Stanford Visualization Group

Polaris (now Tableau)
D3.js
Vega

Jeffrey Heer
@jeffrey_heer

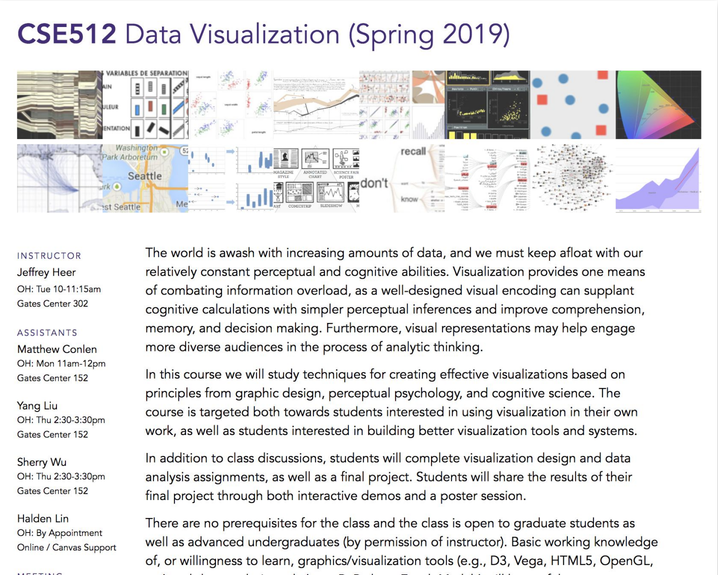
Uncertainty State of the Art

UW Interactive Data Lab

Previously Stanford Visualization Group

Polaris (now Tableau)
D3.js
Vega

Jeffrey Heer
@jeffrey_heer



CSE512 Data Visualization (Spring 2019)

INSTRUCTOR
Jeffrey Heer
OH: Tue 10-11:15am
Gates Center 302

ASSISTANTS
Matthew Conlen
OH: Mon 11am-12pm
Gates Center 152

Yang Liu
OH: Thu 2:30-3:30pm
Gates Center 152

Sherry Wu
OH: Thu 2:30-3:30pm
Gates Center 152

Halden Lin
OH: By Appointment
Online / Canvas Support

The world is awash with increasing amounts of data, and we must keep afloat with our relatively constant perceptual and cognitive abilities. Visualization provides one means of combating information overload, as a well-designed visual encoding can supplant cognitive calculations with simpler perceptual inferences and improve comprehension, memory, and decision making. Furthermore, visual representations may help engage more diverse audiences in the process of analytic thinking.

In this course we will study techniques for creating effective visualizations based on principles from graphic design, perceptual psychology, and cognitive science. The course is targeted both towards students interested in using visualization in their own work, as well as students interested in building better visualization tools and systems.

In addition to class discussions, students will complete visualization design and data analysis assignments, as well as a final project. Students will share the results of their final project through both interactive demos and a poster session.

There are no prerequisites for the class and the class is open to graduate students as well as advanced undergraduates (by permission of instructor). Basic working knowledge of, or willingness to learn, graphics/visualization tools (e.g., D3, Vega, HTML5, OpenGL,

data visualization
lisboa #vislis


Community
Chart Clinic

Thank you!


Welcome
Critique
Visualizing Uncertainty
Community Chart Clinic
End of Event