

Uncertainty for The Novice Fuzz: A Visualization Toolkit

Abigail Dake
Eylul Dogru
Justin Grime
Michael Lai
Thomas Lotz

Roadmap

- Our Project
 - Objective
 - Mathematical Background Information
 - Design and Implementation Process
 - Code Structure
 - User Study Review
- Demo and Results

Project Objectives

- State of the USA
 - Nonpartisan data
 - Widely available
 - Intended to inform decision making
- Larger picture
 - More data available to more people
 - Opportunity for better decision making
 - Info viz can help; Viz uncertainty too for best decisions

Uncertainty for Fuzz

- Numerical uncertainty in statistical data
 - Particularly errors related to sample sizes
- Forecast and uncertainty of forecast
 - Ranking different models
 - Is a forecast meaningful? How meaningful?
- Intended users and what they want to do
- Simple visualizations: Time Series and Bar Chart

About Uncertainty

- Any data and visualization contains some uncertainty due to many reasons:
 - Inaccuracy of instruments
 - Estimations made based on a sample
 - Transformations of the data
 - The visualization itself.
- “Novice” users are often not aware of it.

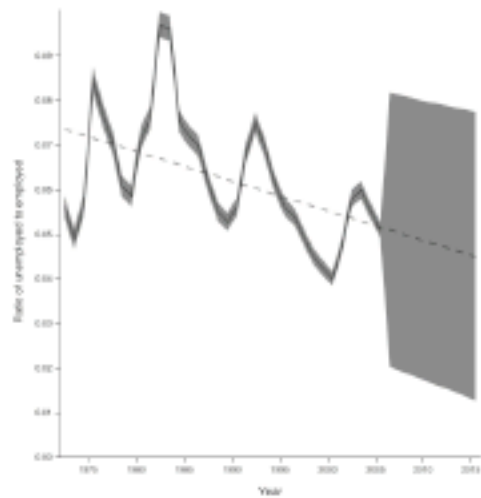
Roadmap

- Our Project
 - Objective
 - **Mathematical Background Information**
 - Design and Implementation Process
 - Code Structure
 - User Study Review
- Demo and Results

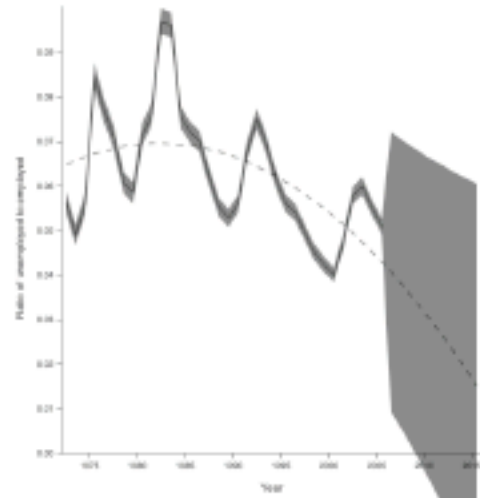
Uncertainty for the Novice

- **Want:** population value
- **Estimate:** from sample
- **Variance** due to limited sample size
 - Wald: $\sqrt{\hat{p}(1 - \hat{p})/n}$
- **Confidence Interval:** region covering the actual (population) value
 - **Confidence:** the chance that the CI covers the actual value
- Other methods (Future Work)

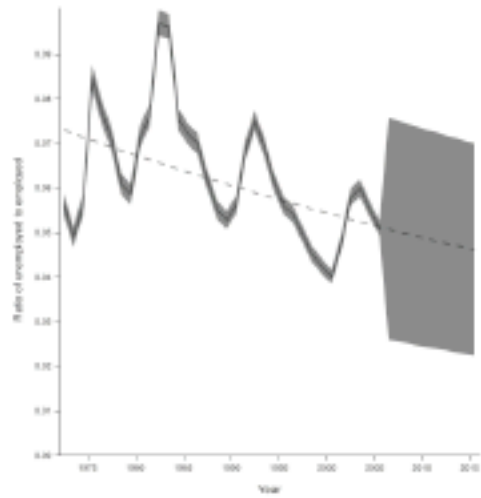
Forecasting



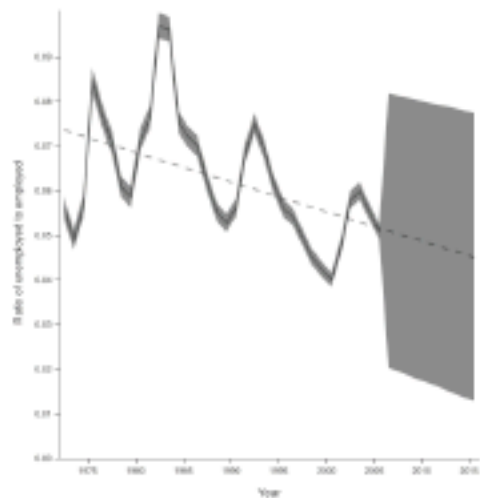
(a) linear



(b) quadratic



(c) exponential



(d) logarithmic

Forecasts

- Regression models:

- Linear, quadratic, exponential, logarithmic

$$y = a + bx + \epsilon$$

$$y = a + bx + cx^2 + \epsilon$$

$$y = ae^{bx} + \epsilon$$

$$y = a + b \ln(x) + \epsilon$$

- Uncertainty of forecast

- Based on regression variance, value uncertainty

- $P(y_0 \in x'_0 \hat{\beta} \pm t_{\alpha/2, n-k-1} s \sqrt{1 + x'_0 (X'X)^{-1} x_0}) \geq \alpha$

- Larger than value uncertainty

Ranking Different Models

- Is a forecast meaningful? How meaningful?
- BIC: Bayesian Information Criterion

$$BIC = n \ln\left(\frac{SS_{err}}{n}\right) + k \ln(n)$$

- R-squared

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$$

$$SS_{err} = \sum_i (y_i - \hat{y}_i)^2$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2$$

Roadmap

- Our Project
 - Objective
 - Mathematical Background Information
 - **Design and Implementation Process**
 - Code Structure
 - User Study Review
- Demo and Results

Design and Implementation Process

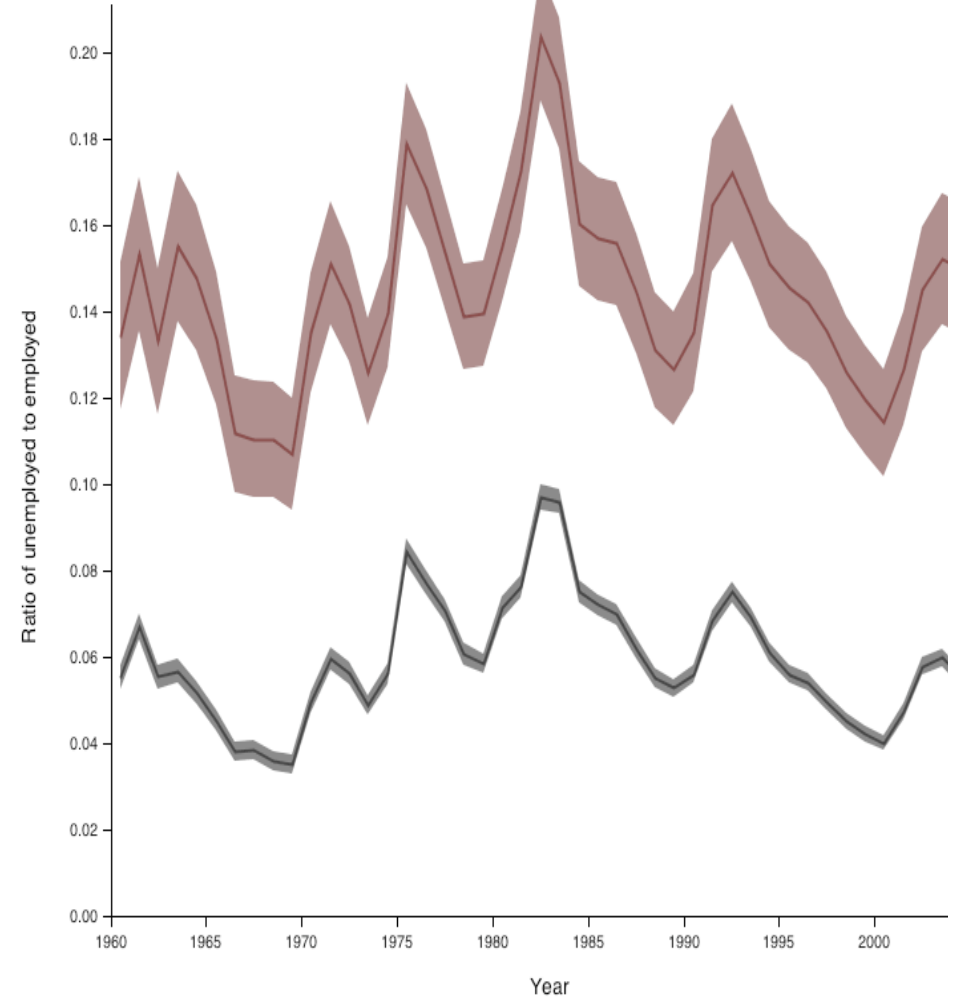
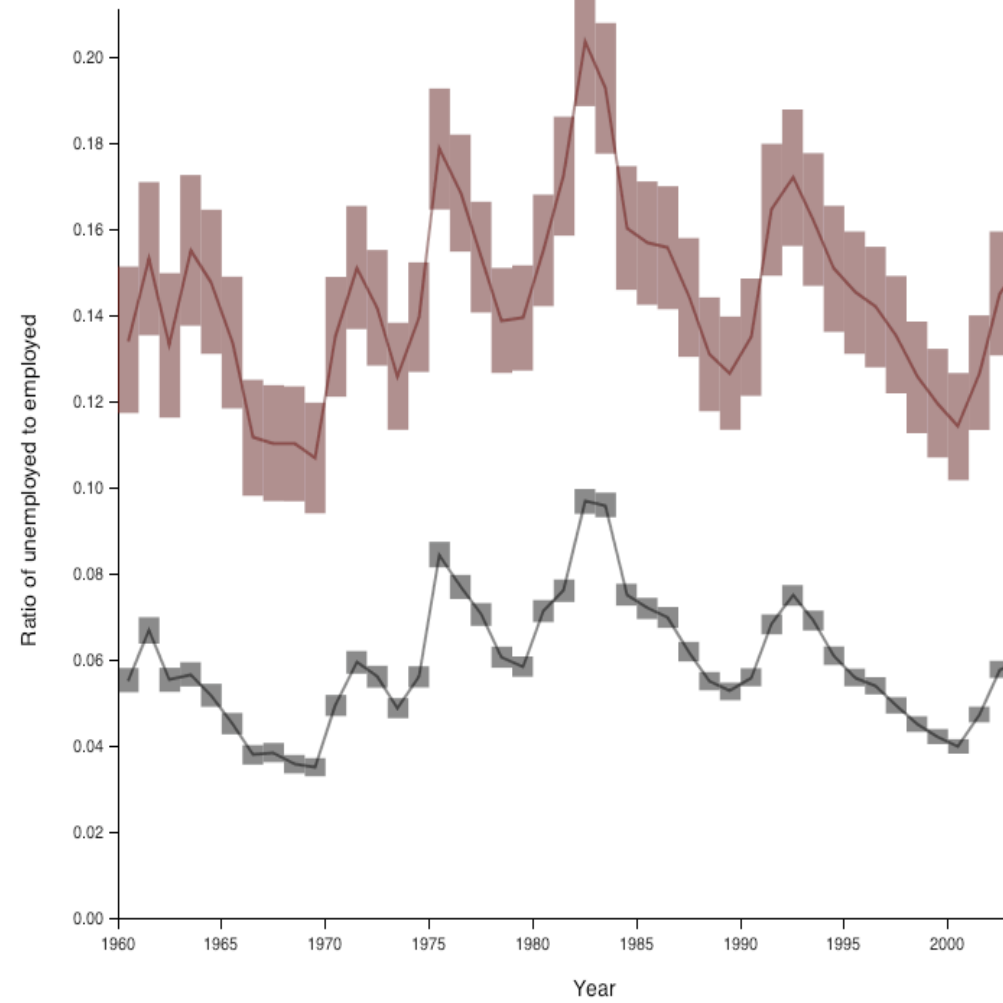
- Participatory Design
 - Interim user
 - Focus: user tasks
 - Focus: eliminating approaches that do not work
- Implementation
- Pilot User Study
 - Can users find correct answers during tasks?
 - User reasoning process and Fuzz design.

Uncertainty Display Principles

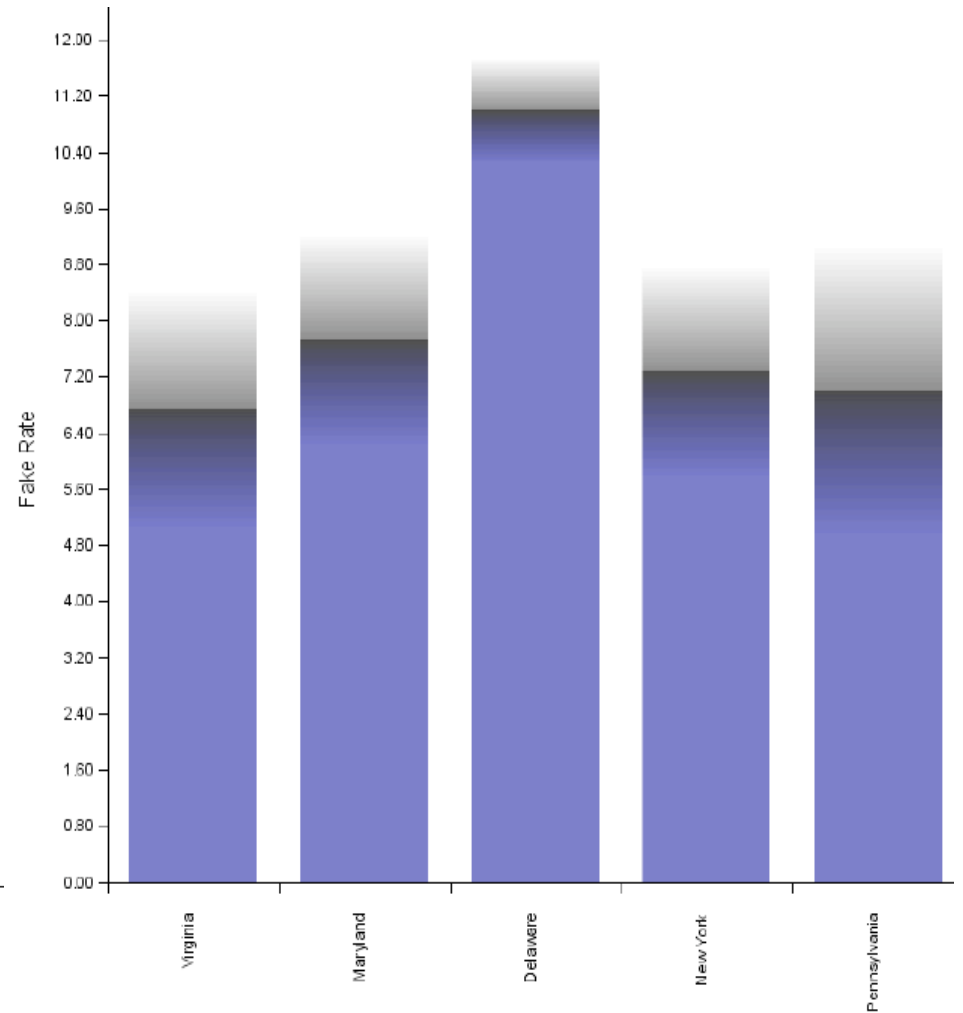
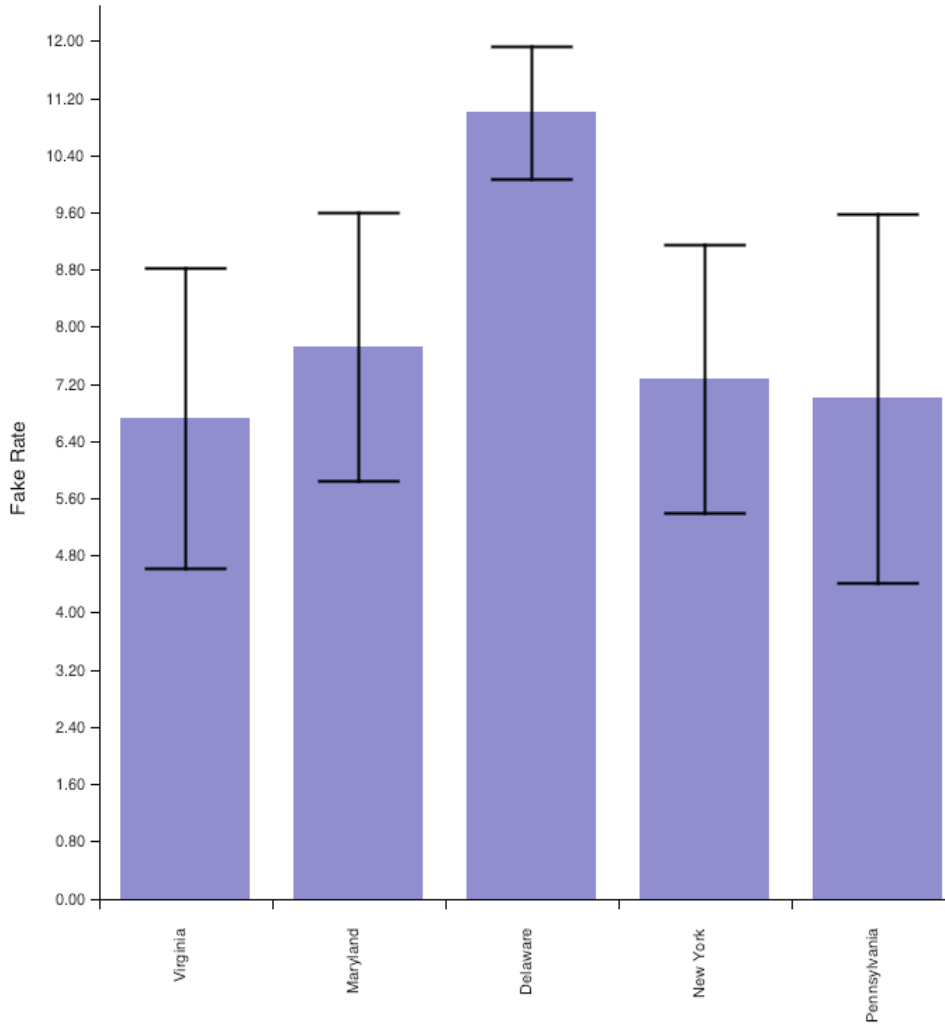
(Cedilnik and Rheingans)

1. Uncertainty should not overshadow the data values.
2. Uncertainty should be “perceptually normalized”, with the same amount of “visual energy” present at every place.
3. Uncertainty representations should be inherently meaningful, having some intuitive way of presenting the uncertainty.
4. The generation of uncertainty visualization should not slow down the visualization.

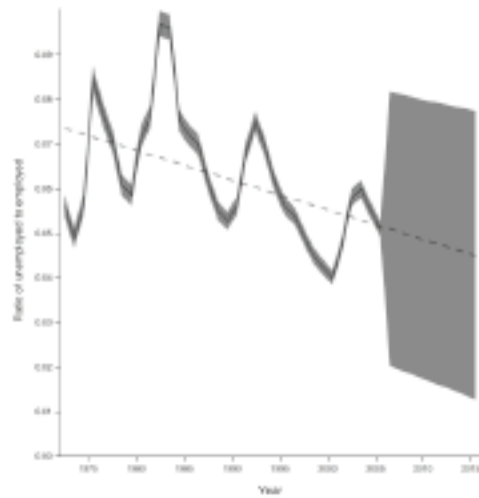
Time Series



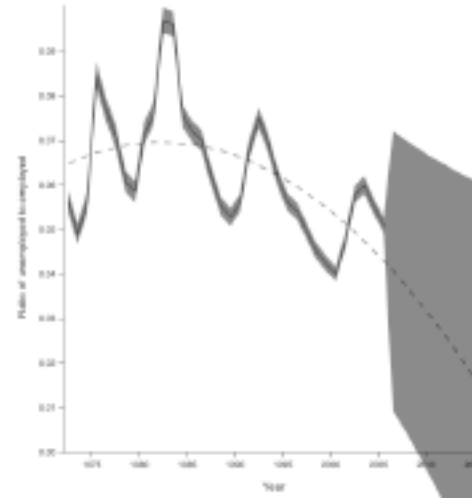
Barcharts



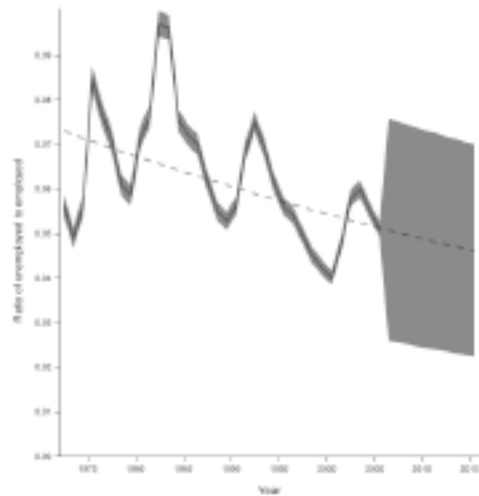
Forecasting



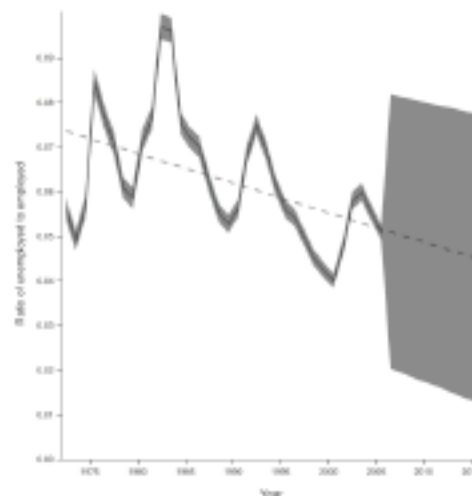
(a) linear



(b) quadratic



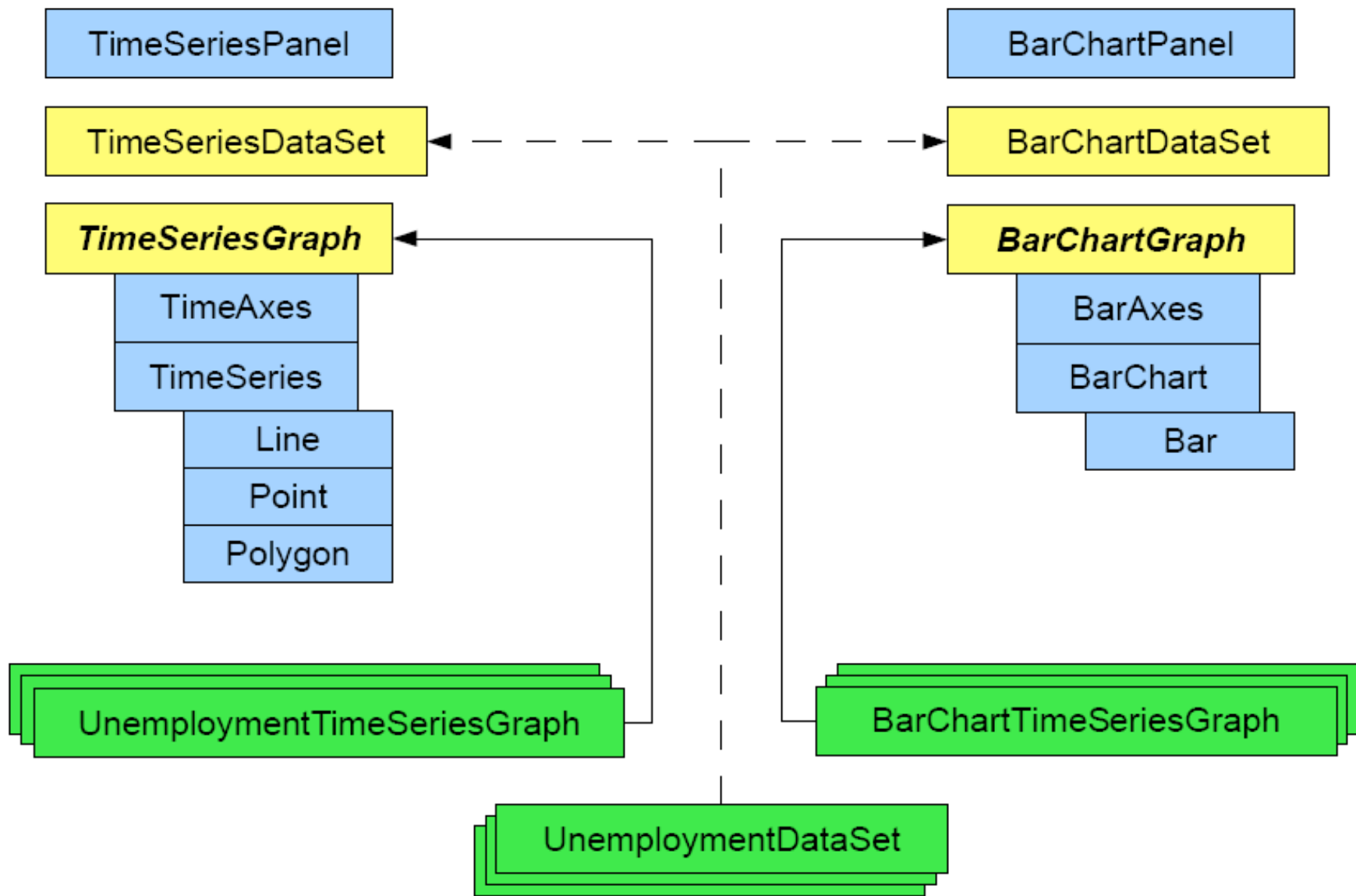
(c) exponential



(d) logarithmic

Code Structure

- Language: Java
- Libraries Used:
 - Visual:
 - AWT/Swing,
 - Piccolo
 - Prefuse
 - Statistics:
 - Jama
 - JSC



Key:

class

interface

abstract class

specializations

User Study Review

- Pilot usability test
- 5 users (graduate students from public policy, computer science and info. science)
- Tested only time series visualization, with unemployment data from Bureau of Labor Statistics
- Task based analysis (4 tasks)
- Short Questionnaire about interface and features

User Evaluation Results

- Almost all users successfully completed all four tasks.
- Results from questionnaire weren't extremely useful.
- Participants went beyond answering task questions
- Uncertainty bars vs Uncertainty regions. (2/3 split inconclusive)
- Proper terminology and layout essential
- Forgetfulness of settings, defaults matter
- Graph on separate axes - Difficult to find but useful
- Zooming feature problematic

Future Work

- User Study
- Statistics:
 - Better Forecasting
 - Confidence Interval calculation
 - Additional models
- Features
 - Easier importing of XY data for scattergram
 - View/download chart data, pdf output

?