Analysis of Global Terrorism Database

Bayesian Inference and Imputing Missing Data

Agenda

- Description of Dataset and Exploratory Data Analysis
- Bayesian Inference:
  Comparison of Terror Incidents in India and Pakistan in 2016
- Missing Data:
  Imputing the Number of Bombings in 1993

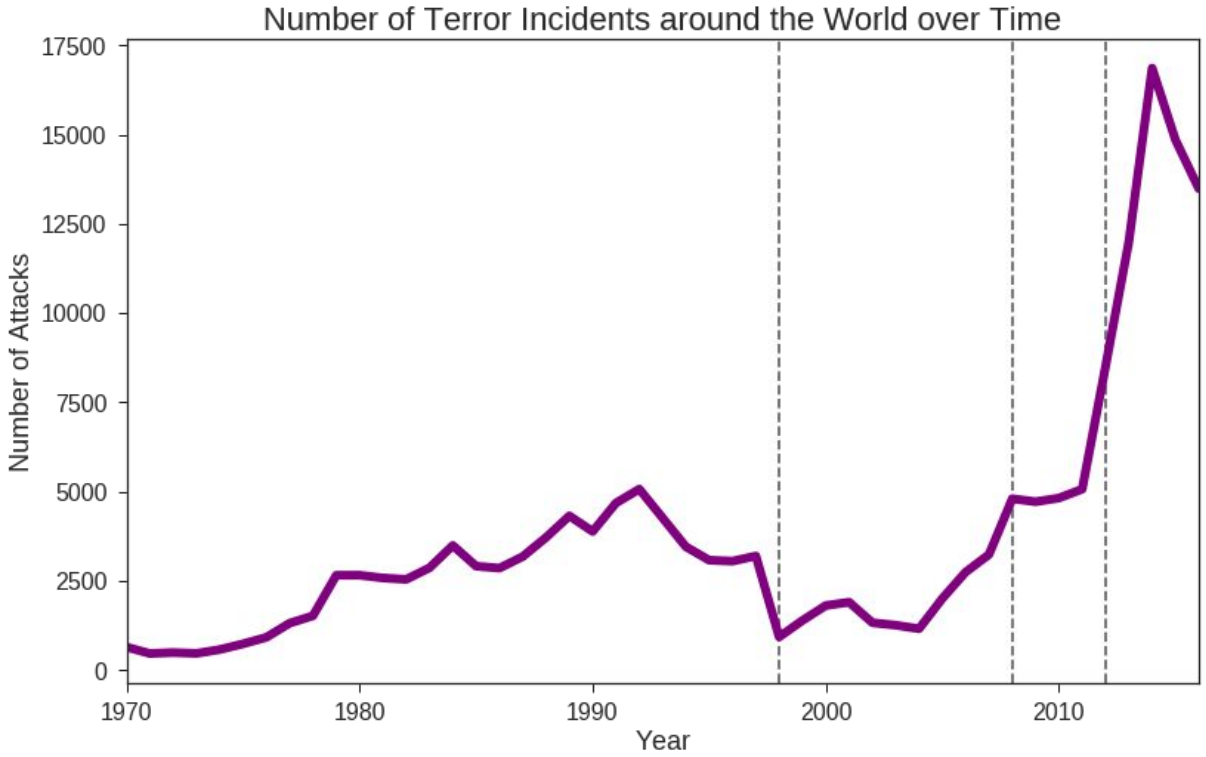
Agenda

- **Description of Dataset and Exploratory Data Analysis**
- Bayesian Inference:
  Comparison of Terror Incidents in India and Pakistan in 2016
- Missing Data:
  Imputing the Number of Bombings in 1993

Description of Global Terrorism Database (GTD)

- Open-source database including information on terrorist attacks around the world from 1970 through 2016 (1993 is missing)
- Currently include more than 170,000 cases
- Maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland
- Variables: >100 variables on location, tactics, perpetrators, targets, and outcomes
- Sources: Unclassified media articles

Exploratory Data Analysis



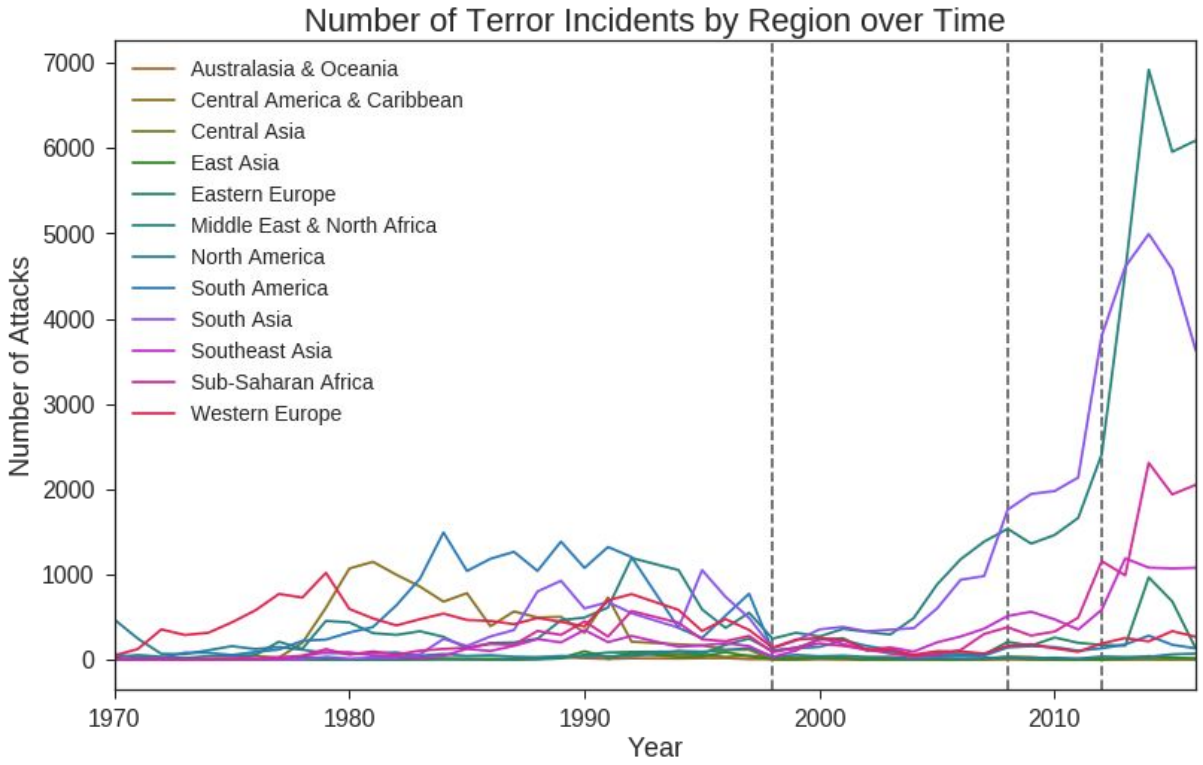
Changes in data collection methodologies occurred around:

- January 1, 1998,
- April 1, 2008
- January 1, 2012

These changes are likely to have an effect on the data collected; proceed with caution when studying trends across these time periods.

http://start.umd.edu/gtd/faq/#q10

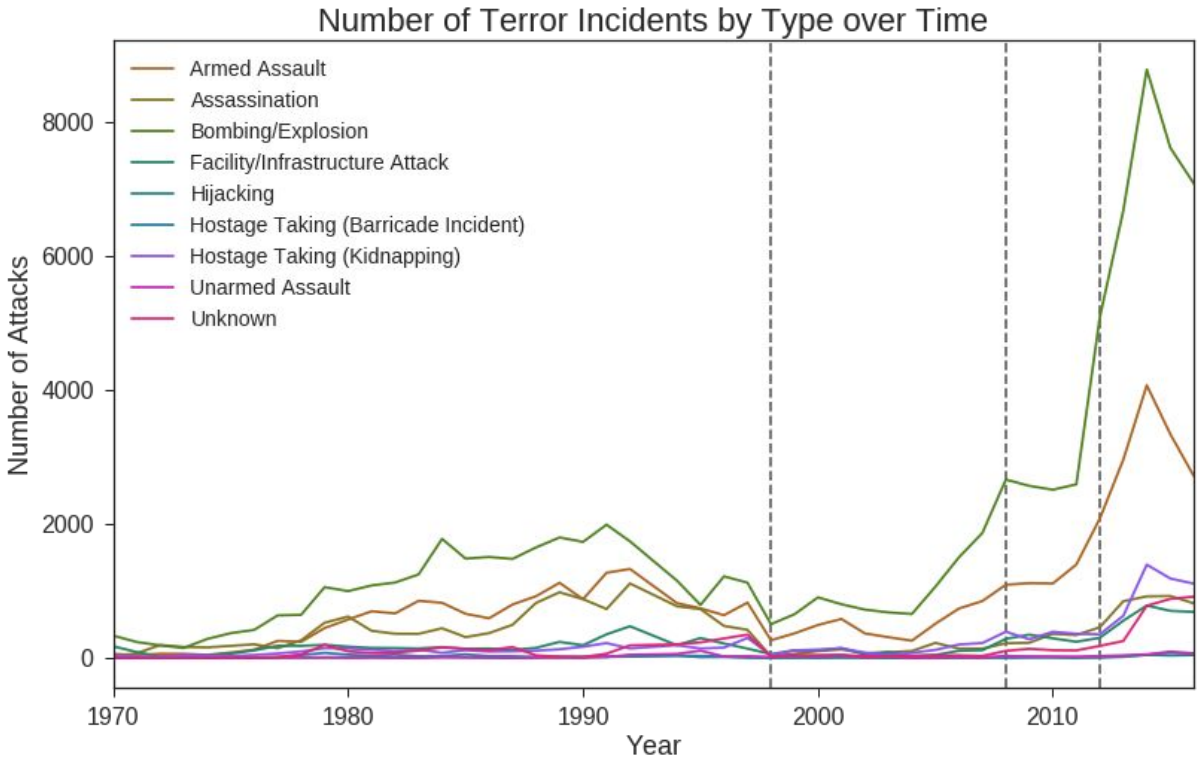
Exploratory Data Analysis



Regions with most terror incidents in recent years are:

- Middle East & North Africa
- South Asia
- Sub-Saharan Africa

Exploratory Data Analysis



Most common types of terror incidents are:

- Bombing/Explosion
- Armed Assault

Agenda

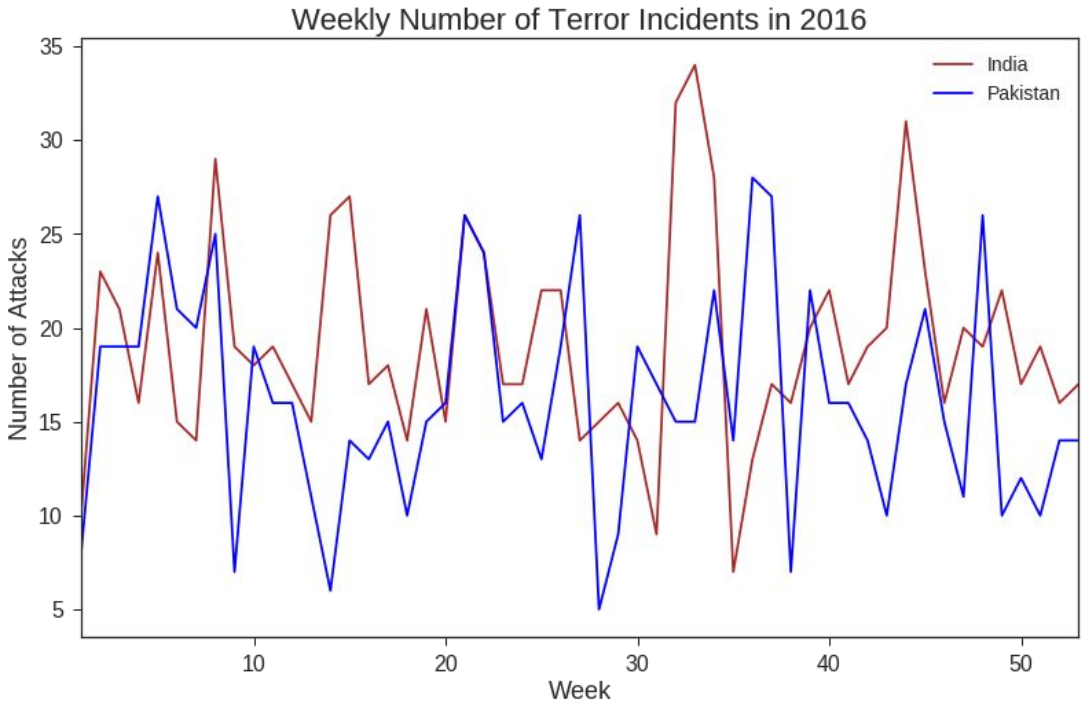
- Description of Dataset and Exploratory Data Analysis
- **Bayesian Inference:**
  Comparison of Terror Incidents in India and Pakistan in 2016
- Missing Data:
  Imputing the Number of Bombings in 1993

Comparison of India and Pakistan in 2016

Total number of terror incidents in 2016:

- India 1019
- Pakistan 861

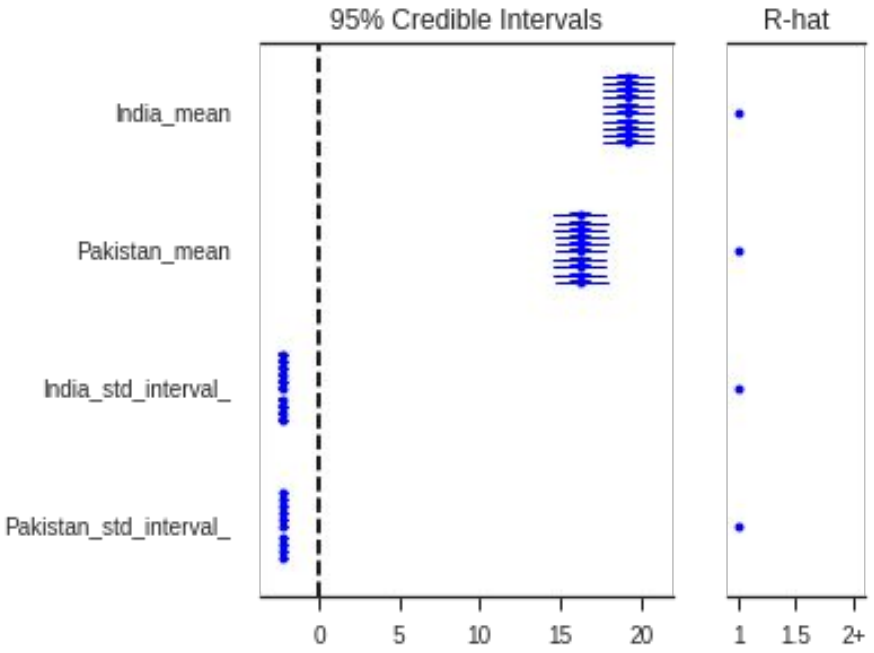
Is the difference statistically significant?



Bayesian Statistics

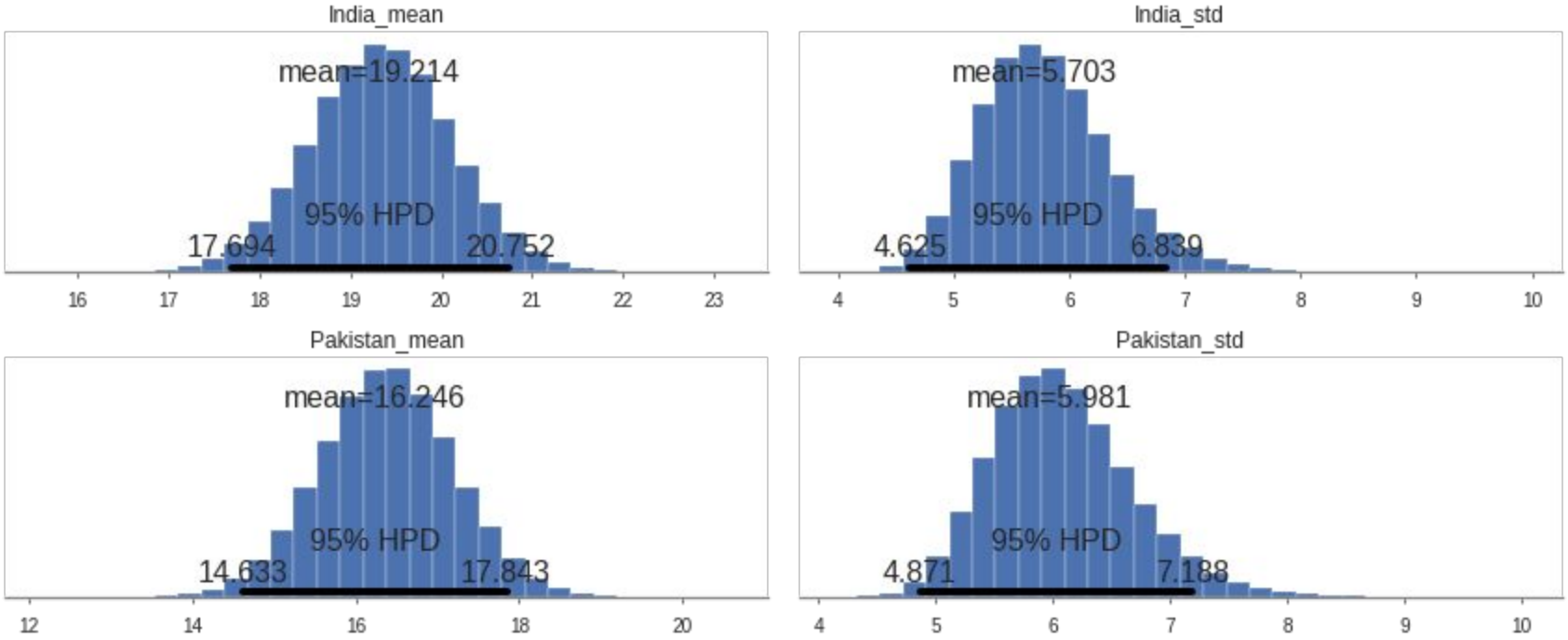
- Priors:
  - Assume normal distribution for the mean
    - Use distributions for the whole region of South Asia (mean, standard deviation)
  - Assume uniform distribution for the standard deviations
- Likelihood:
  - Assume normal distribution
    - Mean is prior distribution of the mean for each group
    - Standard deviation is the prior distribution for the standard deviation for each group
  - Given an observed parameter which is the actual data
- Posterior:
  - Update according to prior beliefs and observed data

Bayesian Inference

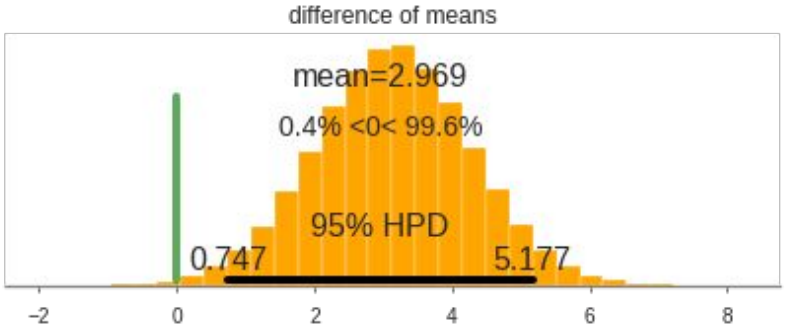


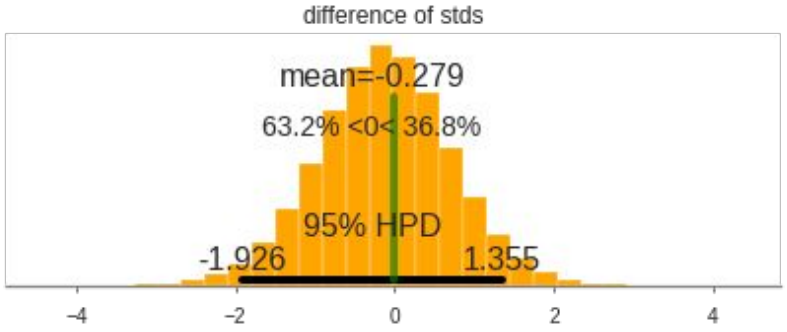
R-hat = 1 indicates convergence has been achieved

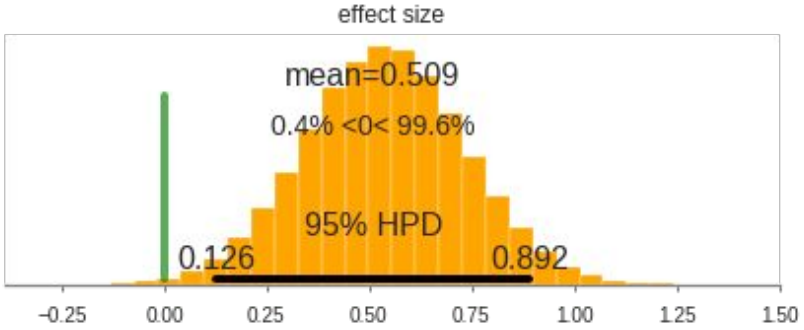
Bayesian Inference



Bayesian Inference





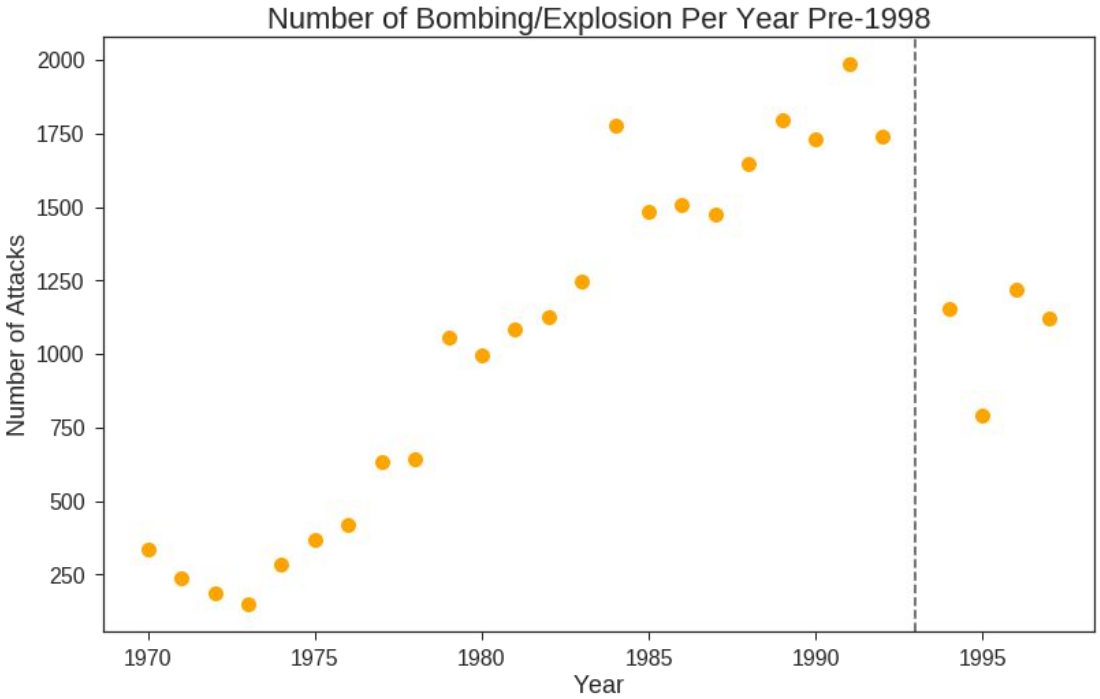


Interpretation:

The difference in the average numbers of attacks per week in India and Pakistan in 2016 is statistically significant.

Agenda

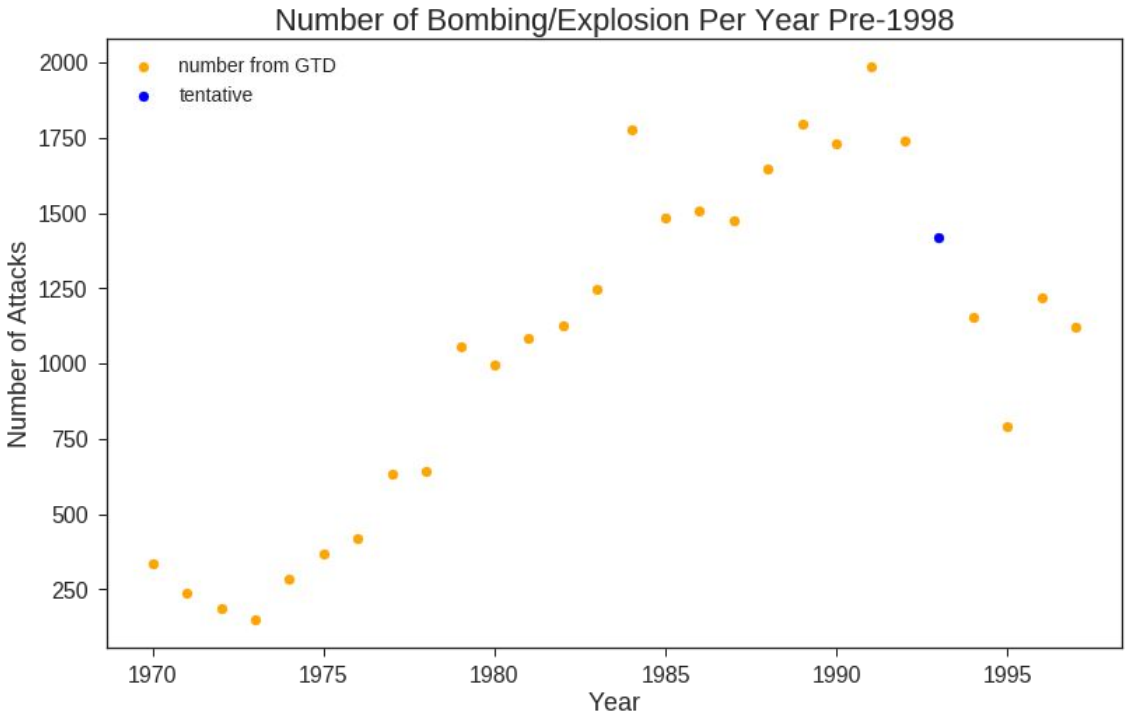
- Description of Dataset and Exploratory Data Analysis
- Bayesian Inference:
  Comparison of Terror Incidents in India and Pakistan in 2016
- **Missing Data:**
  Imputing the Number of Bombings in 1993



There appears to be a downward trend between 1991 and 1995 after the upward trend since the early 1970s.

Impute the number in 1993 by simply averaging the numbers in years 1991, 1992, 1994, and 1995

Estimated number of bombing/explosion in 1993: 1417



Questions?

Is there a methodological reason for the decline in the data between 1997 and 1998, and the increases since 2008 and 2012?

While efforts have been made to assure the continuity of the data from 1970 to the present, users should keep in mind that the data collection was done as events occurred up to 1997, retrospectively between 1998 and 2007, and again concurrently with the events after 2008. This distinction is important because some media sources have since become unavailable, hampering efforts to collect a complete census of terrorist attacks between 1998 and 2007. Moreover, since moving the ongoing collection of the GTD to the University of Maryland in the Spring of 2012, START staff have made significant improvements to the methodology that is used to compile the database. These changes, which are described both in the GTD codebook and in this **START Discussion Point on *The Benefits and Drawbacks of Methodological Advancements in Data Collection and Coding: Insights from the Global Terrorism Database (GTD)***, have improved the comprehensiveness of the database. Thus, users should note that differences in levels of attacks before and after January 1, 1998, before and after April 1, 2008, and before and after January 1, 2012 may be at least partially explained by differences in data collection; and researchers should adjust for these differences when modeling the data.

http://start.umd.edu/gtd/faq/#q10