



PHUSE US Connect 2024

PP12 - Mining Historical Data and Epidemiological Evidence for Real World Studies

Shubhranshu Dutta, University of Rochester

shudutt@gmail.com

Integrating Generative AI Tools To Analyze The Efficacy of Pandemic Response

1. **Model Selection** - Identify which LLM (Large Language Model) - GPT 4, Gemini (Integrated through Bard), BLOOM, LLaMA, Claude, etc. - is to be used for data collection and parameter definition

- Each LLM has its own strengths/weaknesses/abilities (eg. conceptual understanding and representation vs. computational strength)
- Clear parameters for selection can be defined to place importance on capabilities of priority - for example, access to the internet, or a ranked ability to retrieve certain types of information.
- Features of Interest:
 - Multimodality (ability to work with both text and image (Gemini))
 - Access to Internet (Bard)
 - Accuracy and Computational Ability (GPT 3.5 and 4)
 - Reduced Bias/Hallucinations (BLOOM/LLaMA)
 - Pattern Recognition & Adaptive Learning (GPT 4 and Gemini)

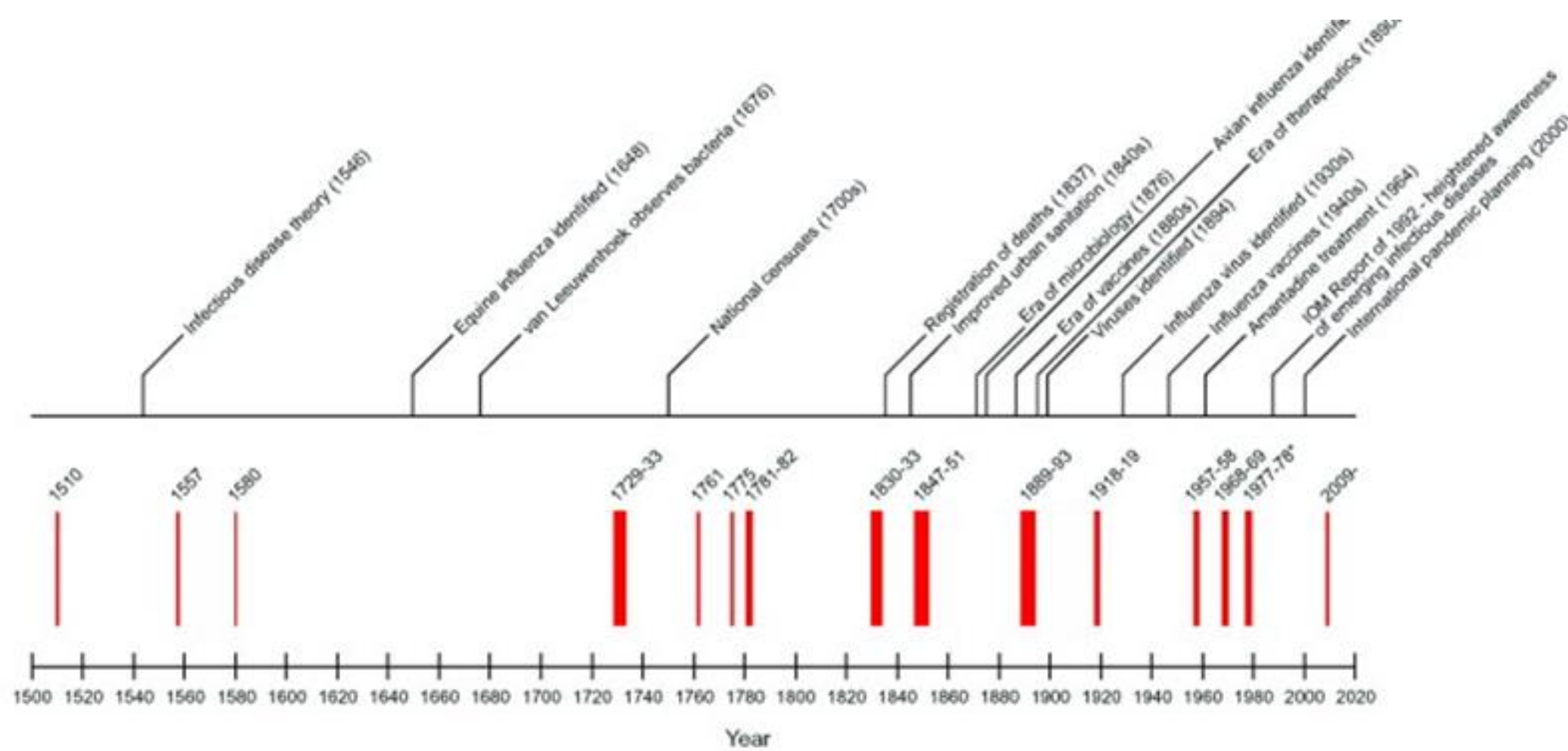
2. **Prompt Engineering** - Employ prompt-engineering techniques to refine searches and increase return efficacy

- Potential Methods:
- Self-Ask Prompting
 - Meta Prompting
 - Chain-of-Thought Prompting
 - ReACT
 - Symbolic Reasoning and PAL
 - Iterative Prompting
 - Sequential Prompting
 - Self-Consistency Method
 - Automatic Reasoning & Tool Use (ART)

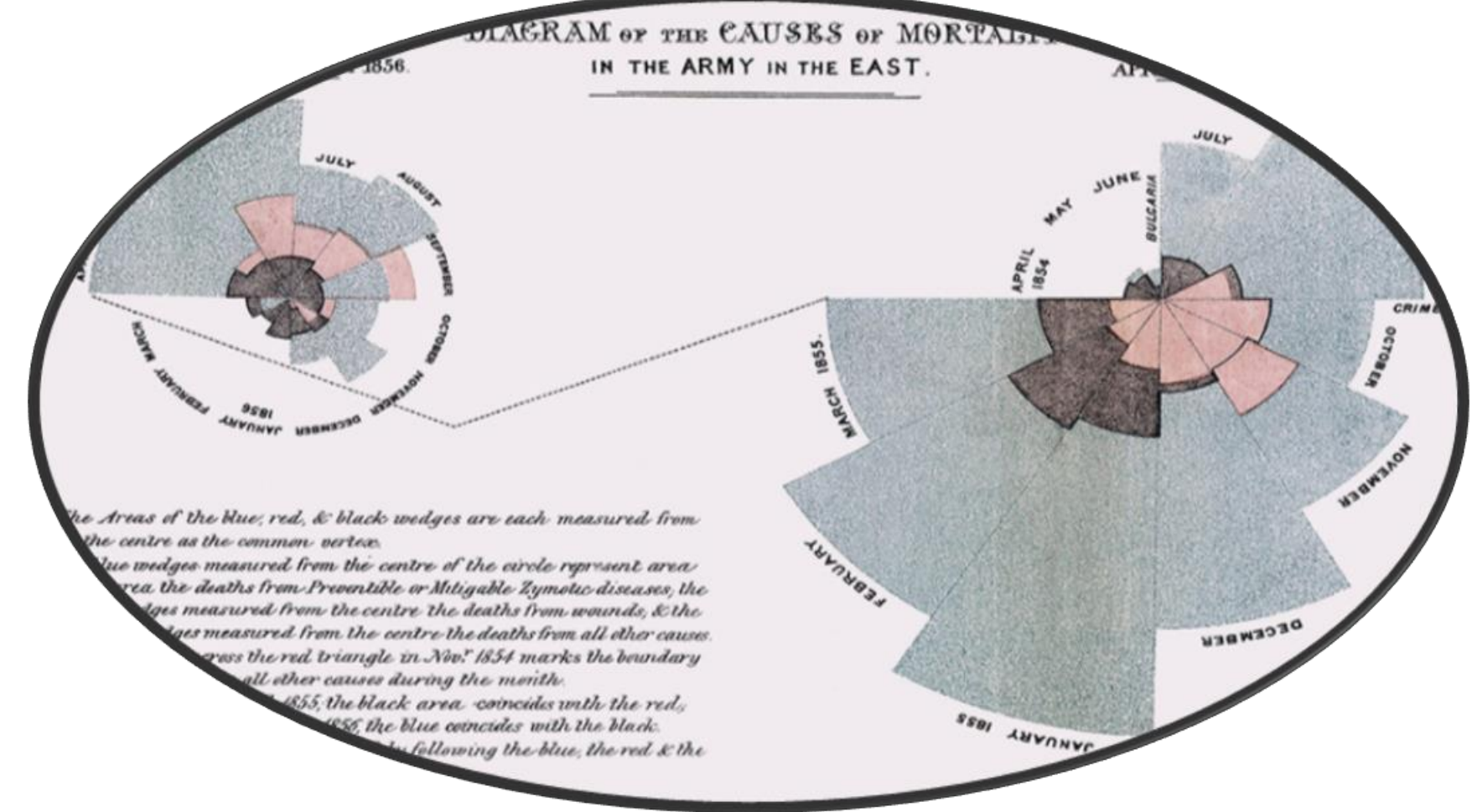
*Ask targeted questions ex: duration, geographic location, phenotypes, genotypes, pathogens, vector mode, etc.

3. **Fine-Tuning Model** - Possibly Fine-Tune LLMs to our own research specifications by preparing and uploading training data (perhaps data collected from prior pandemics/health crises, or for our purpose - a hand-done model analysis of a past pandemic with parameters specified to teach the LLM)

4. **Retrieval Augmented Generation** - Going back to original references and ensuring content validity/accuracy. This may be supplemented by further LLM queries.



Data Visualization: Time Trending: Pandemic Influenza's 500th Anniversary



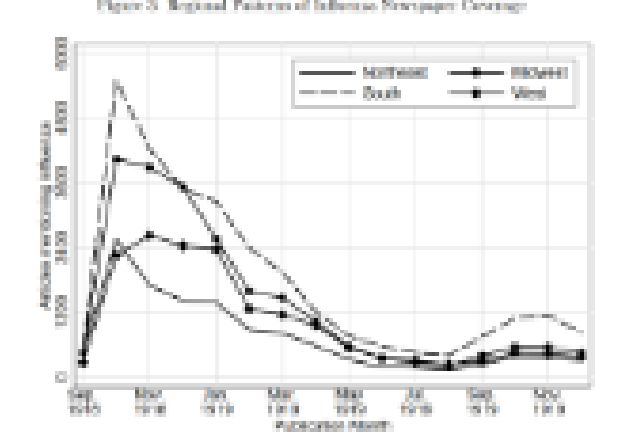
Data Visualization: Florence Nightingale's Rose Chart/Coxcomb graph representing unnecessary deaths during the Crimean War because of preventable infections among wounded soldiers.

Prompt	BARD	GPT 3.5
List of all epidemics in world history	Unfortunately, providing a complete list of all epidemics in world history is virtually impossible, for several reasons: • Lack of comprehensive records: Records of historical outbreaks, especially in ancient times, are often incomplete or inaccurate. Many epidemics may have gone entirely unrecorded. • Varied definitions: Distinguishing between an epidemic and an endemic disease can be subjective. An outbreak in one region might be considered an epidemic, while similar cases in another area might be seen as endemic.	Creating a comprehensive list of all epidemics in world history is an enormous task due to the vast number of outbreaks spanning millennia across different regions. Here's an attempt to provide a broad overview: Ancient Times: • Unknown BCE Epidemics: Unrecorded epidemics likely occurred in ancient civilizations but lack detailed historical documentation.
	• Continual occurrence: New epidemics and resurgences of old ones continuously occur, making the list ever-changing. Therefore, instead of a complete list, I can offer several alternatives to explore historical epidemics: By Time Period: • Ancient Epidemics: - o Plague of Athens (430 BC) - possibly typhoid fever - o Antonine Plague (165-180 AD) - possibly smallpox - o Plague of Cyprian (249-262 AD) - possibly smallpox • Medieval Epidemics: - o Justinian Plague (541-549 AD) - bubonic plague - o Black Death (1347-1353) - bubonic plague - o Great Bubonic Plague of London (1665) - bubonic plague • Modern Epidemics: - o Spanish Flu (1918-1919) - influenza A virus - o Asian Flu (1957-1958) - influenza A virus - o AIDS epidemic (ongoing) - HIV/AIDS - o COVID-19 pandemic (ongoing) - SARS-CoV-2 virus By Region: • European Epidemics: - o Dancing Plague of Strasbourg (1518) - likely mass psychogenic illness - o Great Famine of Ireland (1845-1852) - potato blight	Classical Period and Middle Ages: • Plagues of Antiquity: This includes plagues like the Plague of Athens, Antonine Plague, and Justinian Plague, impacting Mediterranean civilizations. • Plague of Cyprian (250-270 AD): A pandemic affecting the Roman Empire. Renaissance and Early Modern Period: • Black Death (1347-1351): One of the most devastating pandemics in history, caused by the Bubonic Plague in Europe. • Sweating Sickness (15th-16th centuries): An enigmatic disease that affected England. Industrial Revolution to Early 20th Century: • Cholera Pandemics: Occurred in multiple waves across continents in the 19th century. • 19th-century Smallpox Epidemics: Significant outbreaks before widespread vaccination efforts.

History of RWD in pandemics.	The history of real-world data (RWD) in pandemics is a fascinating and evolving story, one that goes beyond simply pinpointing a single "first" instance. It's more about understanding the gradual development of data collection practices and their increasing sophistication over time. Let's delve deeper into this remarkable history: Early Stages (Pre-formalized RWD): • Ancient Pandemics (Plague of Athens, Justinian Plague): While formal data collection wasn't prevalent, these pandemics left traces of information in the form of chronicles and records. These might mention mortality rates, geographical spread, or basic symptom descriptions. While not structured by today's standards, these glimpses provide valuable insights into early understandings of pandemics. • Medieval Pandemics (Black Death): Compared to ancient times, data collection became more elaborate. Chronicles tracked affected areas, demographics, and even anecdotal accounts of symptoms and treatments. This offered richer information for understanding the pandemic's impact, even if limited by the times.	Real-world data (RWD) has been crucial in understanding, managing, and responding to pandemics throughout history. Here's a brief overview of RWD in different pandemics: Spanish Flu (1918-1919): • Data Collection: Epidemiologists collected RWD on infection rates, mortality, hospitalizations, geographic spread, and demographics to understand the impact and spread of the virus. HIV/AIDS Pandemic (1980s-Present): • Longitudinal Studies: RWD from longitudinal studies provided insights into disease progression, treatment effectiveness, and the impact of HIV/AIDS on different populations over time.
------------------------------	---	---

References:

1. Florence Nightingale, data visualization visionary - JSTOR DAILY. (n.d.). <https://daily.jstor.org/florence-nightingale-data-visualization-visionary/>
2. Morens DM, Taubenberger JK, Folkers GK, Fauci AS. Pandemic influenza's 500th anniversary. Clin Infect Dis. 2010 Dec 15;51(12):1442-4. doi: 10.1086/657429. Epub 2010 Nov 10. PMID: 21067353; PMCID: PMC3106245.

Spanish Flu RWD data collection and analysis publications with citations	Spanish Flu RWD Data Collection and Analysis: A Look at Key Publications with Citations While the term "real-world data" (RWD) wasn't used during the Spanish Flu pandemic, the era witnessed significant efforts to collect and analyze data relevant to the outbreak. Here are some key publications that exemplify RWD collection and analysis practices of the time: Mortality Data: • John Wilson, "Influenza: A Study of the History, Epidemiology, Pathology and Bacteriology" (1919): This comprehensive work analyzes global mortality data, providing insights into age-specific trends, geographic spread, and potential risk factors. It remains a valuable resource for understanding the pandemic's impact. • The Rockefeller Foundation's Epidemiological Studies: The Rockefeller Foundation conducted extensive studies after the pandemic, collecting mortality data from various countries. These studies examined death rates, identified high-risk groups, and investigated the influence of public health measures. Spatial Analysis: • Frederick C. Croxton, "Influenza of 1918: Statistics of the Epidemic as Recorded in the Weekly Reports of the Health Officers of the 88 Cities of the United States" (1922): This publication	Research on the Spanish Flu, also known as the 1918 influenza pandemic, has been extensive, offering various perspectives on its impact, epidemiology, and historical significance. The availability of real-world data (RWD) for this historical event is limited, given the era's data collection methods. However, numerous studies have analyzed available historical records, including mortality data, medical reports, and archival materials. Here are some key publications related to the Spanish Flu with citations: 1. Barry, John M. "The Great Influenza: The Story of the Deadliest Pandemic in History." Penguin Books, 2005. • This book provides a comprehensive narrative of the Spanish Flu, detailing its origins, global spread, societal impacts, and the response of medical professionals
Role of statistical analysis in controlling the spread of Spanish flu in usa	The role of statistical analysis in controlling the spread of the Spanish Flu in the USA (1918-1919) was complex and limited due to various factors: Limited Data and Collection Methods: • Data collection systems for cases and deaths were inconsistent and incomplete across regions, making accurate trend analysis and comparisons difficult. • Reporting varied by age groups, cause of death definitions, and reporting formats, further hindering data standardization and analysis. Challenges in Analysis and Interpretation: • The novelty of the virus and limited understanding of its transmission made it challenging to interpret data and draw direct correlations between interventions and outcomes. • Statistical techniques of the time were less sophisticated compared to today, limiting the ability to model complex relationships and analyze large datasets. • Social and economic factors impacting the pandemic (e.g., wartime mobilization, population movements) complicated isolating the exact impact of statistical analysis on public health responses.	During the Spanish Flu pandemic in the early 20th century, statistical analysis played a crucial role in understanding and attempting to control the spread of the virus in the USA. While the methods and technology were less advanced compared to today, statistical analysis contributed significantly in several ways: 1. Epidemiological Surveillance: Statisticians and public health officials used basic epidemiological principles to track the spread of the virus. They collected and analyzed data on illness and death rates to identify patterns, affected demographics, and geographic areas experiencing higher infection rates. 2. Mortality Rates Analysis: Statisticians calculated and analyzed mortality rates, comparing them across regions, age groups, and socioeconomic factors. This analysis
Research Articles on Spanish Flu Data Visualization , Graphs, and Mortality Rates in the USA:	Finding research articles directly visualizing Spanish Flu data in the USA presents challenges due to data limitations and historical context. However, several articles offer valuable insights and alternative approaches: Articles on Challenges and Opportunities: • "Data Stories of the Great Influenza Pandemic" by the U.S. National Archives: This article explores the difficulties and opportunities of visualizing Spanish Flu data in the USA, showcasing historical maps and charts with discussions on their interpretations and limitations. • "Mapping the Flu: Lessons from the 1918 Pandemic" by The New York Times: This interactive article examines data limitations while showcasing potential visualization approaches by overlaying historical maps with contemporary data.  • "Visualizing Historical Public Health Data: From Paper Forms to Interactive Maps" by the Journal of Public Health Management and Practice: This article	I can't directly access real-time databases or specific articles. However, here's a generalized approach to find research articles on Spanish Flu data visualization, graphs, and mortality rates in the USA: 1. Database Search: Utilize academic databases like PubMed, JSTOR, or Google Scholar. These platforms compile a wide range of scholarly articles. 2. Keywords: Use specific keywords such as "Spanish Flu," "1918 influenza pandemic," "mortality rates," "data visualization," "graphs," and "United States" in your search. 3. Search Phrases: Combine these keywords in search phrases like "Spanish Flu mortality rates visualization USA" or "graphs of 1918 influenza pandemic in the United States" to refine your results. 4. Filter and Sort: After getting search results, use filters to narrow down by publication date, article type, or specific aspects you're
Role of statistical analysis in controlling the spread of Spanish flu in USA	The role of statistical analysis in controlling the spread of the Spanish Flu in the USA (1918-1919) was complex and limited due to various factors: Limited Data and Collection Methods: • Data collection systems for cases and deaths were inconsistent and incomplete across regions, making accurate trend analysis and comparisons difficult. • Reporting varied by age groups, cause of death definitions, and reporting formats, further hindering data standardization and analysis. • Real-time data availability was limited, often relying on weekly or monthly reports, restricting timely interventions based on data insights.	During the Spanish Flu pandemic in the early 20th century, statistical analysis played a crucial role in understanding and attempting to control the spread of the virus in the USA. While the methods and technology were less advanced compared to today, statistical analysis contributed significantly in several ways: 1. Epidemiological Surveillance: Statisticians and public health officials used basic epidemiological principles to track the spread of the virus.