UNVEILING THE HIDDEN ENVIRONMENTAL NARRATIVE OF NEW YORK STATE: A DATA VISUALIZATION JOURNEY

DVIZ, Herbstsemester 2023

Teresa WIndlin

TABLE OF CONTENTS

INTRODUCTION

This project delves into the often-overlooked phenomenon of land-based material leaks in New York State, shedding light on their environmental impact.

MOTIVATION

The motivation for choosing this topic stems from the relative lack of public discourse on land spills compared to aquatic spills. While water-based leaks are well-documented, their land-based counterparts, despite being equally harmful, receive less attention. This project aims to fill that gap.

New York State was chosen due to its comprehensive dataset, which offers valuable insights into the causes of spills and the effectiveness of clean-up efforts (NY Open Data, 2024)

GOAL

The primary goal is to raise awareness through an easy-to-read report **among** environmentally conscious **residents of New York State. The project also sets the stage for future actions to reduce spills through prevention, like analyzing and addressing common spill causes, and improving response strategies based on past** clean-up **records.**
The findings are designed to be published on a govornment news page, potentially distributed to local news outlets and schools.

The intended message of this project is to highlight that, although overall spill incidents are decreasing in New York State, larger spills contribute most to the total volume. This points to a need for targeted action in densely populated areas and industrial zones, as they are most effected. Additionally, the project includes a Q&A section on how to respond to a spill discovery.

VISUALISATION OF SPILLS: LITERATURE

First, we explore various visualizations and literature related to spills, focusing on New York State's land-based incidents compared to aquatic spills. While large oil spills at sea, like those detailed in the ITOPF report (ITOPF, 2021) , receive significant attention, land-based spills, the largest often being events such as train derailments (Llamas, 2023), are less highlighted.
Data visualization tools are crucial in conveying the extent of these spills. (Etkin, 2001) presents a graph showing the annual amount of oil spilled in the U.S. by source, and (Hernandez, 2020) utilizes satellite imagery overlaid with spill information, providing a clear geographical perspective.

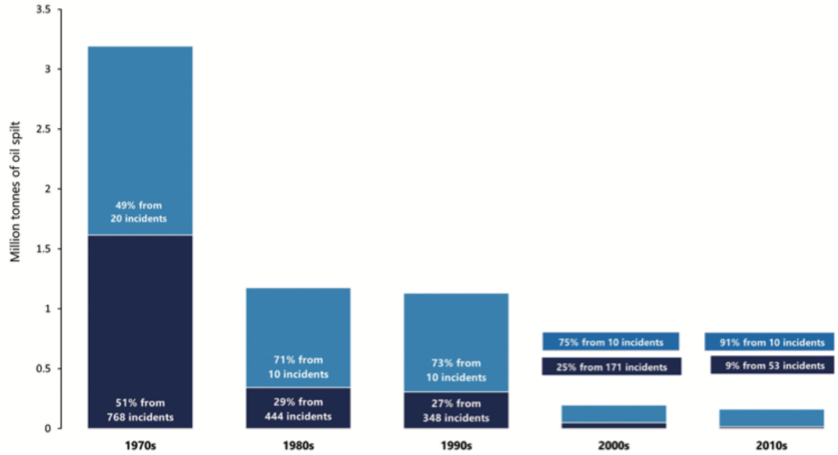

FIGURE 1: SPILLS 7 TONNES AND OVER PER DECADE SHOWING THE INFLUENCE OF A RELATIVELY SMALL NUMBER OF COMPARATIVELY LARGE SPILLS ON THE OVERALL FIGURE (ETKIN, 2001)

FIGURE 2: SATELITE IMAGE OVERLAYED WITH SPILL AREA (HERNANDEZ, 2020)

DATA PREPARATION

fIn preparing the data for this report, I refined our dataset by removing incomplete or incorrect entries, such as records with incorrect county information or missing spill quantities. All measurements were standardized to the metric system for consistency.

The analysis focuses on data from 1990 onwards to provide a clear understanding of recent spill trends. Additional details like geographical coordinates and cleanup durations were incorporated to enhance the dataset's depth and usefulness.

DESIGN CHOICES

The design choices for the report were made with the intent of making the information accessible and engaging. The report is designed for desktop usage, with a layout that presents text and graphs side by side for easy comparison and understanding.
In terms of color usage, the report employs purposefully colorful visuals. This approach aims to make the serious and often tragic topic of spills more approachable and engaging for the reader. The idea is to draw attention to the issue without overwhelming the audience with a somber tone.

Hazardous Material	Other	Oxygenates	Petroleum

The graphs themselves are designed to be clean and clear, using a white background. This choice aligns with the report's intended publication on the official website of New York State, where a professional and uncluttered appearance is essential. The simplicity of the graphs ensures that the focus remains on the data and its implications, facilitating better understanding and engagement from the audience.

DATA VISUALISATIONS

DECADES OF SPILLS RECORDED: A DETAILED LOOK

- **Initial Query**: Analyzing the evolution of spill frequency and material composition since 1990.
- **Message**: Small petroleum spills occurred consistently, with a decrease in overall incidents after the 2000s, indicating improvement in processes.
- **Challenge**: Displaying trends for both common and rare spill materials clearly. Visually distinguishing between frequency (this chart) and quantity (next charts). Caused most confusion based on feedback.
- **Chart Options**:
    - **Stacked (Area) Boxplot**: User-friendly, good for overarching trends, but insufficient detail for rarer spills.
    - **Waffle Chart**: Visually impactful for representing specific numbers of spills, making it ideal for conveying the count of incidents. Yet challenging in depicting gradual trends for all material groups without over or underrepresenting data.

    - **Yearly Scrollbar Visualization**: Detailed, individual year data, but requires user action, hence less intuitive for observing long-term trends.

- **Chosen Chart & Detail**: Selected a boxplot to depict each year with distinct boxes per material type, colored differently. The design included a trendline to signify the overall decrease and annotations to stress the frequency of minor petroleum spills.

QUANTIFYING SPILL IMPACTS OVER TIME

- **Initial Query**: Measuring the volume impact of minor, large and major spills of different material families.

- **Message**: Major spills account for most of total volumes. The largest spill (raw sewage) is omitted from subsequent analyses due to its data skewing effect and comparatively minimal environmental harm.

- **Challenge**: Illustrating complex data layers (volume category, total quantity, material) and handling outliers.

- **Chart Options**:

    - **Lorenz Curve**: Effectively shows disproportionate impact of outlier and major spills but isn't user-friendly for casual viewers.

    - **Bubble Chart (by time-period)**: Intuitively represents spill volume and material type through bubble size and color. When showed to other people led to confusion because different circle for every time-period and different colors.

- **Chosen Chart & Detail**: Bubble Chart (without time) representing volume categories and materials, with size indicating total volume spilled. Material families are doubly represented by placement and color. Focusing on key message.

GEOSPATIAL ANALYSIS OF SPILLS: IMPACTFUL COUNTIES HIGHLIGHTED

- **Initial Query**: Visualizing spill volumes and materials by counties of New York State.

- **Message**: Counties with highest total quantity all around New York City, a densly populated area and the counties with most hazardous spills also close to each other in an industrial area.

- **Challenge**: Combining detail and clarity in the map and deciding between interactive or static displays.

- **Chart Options**:

    - **Map with circle for counties**: Simple visual using circle size for quantity and color for most common material spilt. Lacking in useful detail.

    - **County-wise Pie Charts**: Detailed with info on material families yet overwhelming for viewers.

    - **Interactive vs. Non-Interactive Maps**: Interactive maps enhance user engagement but may add complexity.

- **Chosen Chart & Detail**: A combination map displaying spill quantity and dominant material per county. Special focus on counties with the highest total or most hazardous spills. Design included color coding to differentiate regions based on spill characteristics. The interactive format was chosen to allow detailed exploration by users, providing a more engaging and informative experience. Data was processed to highlight specific counties, ensuring the map remains uncluttered while still delivering key insights.

SUNBURST CHART: "REASON BREAKDOWN IN SPILLS"

- **Initial Query**: Delineating the variety of materials involved in spills across regions.
- **Message**: The chart reveals the diversity of spill causes, with specific materials associated more with either industrial or residential zones.
- **Challenge**: Creating a hierarchical representation of material data that is informative and not overwhelming.
- **Chart Options**:
    - **Nested Pie Charts**: Could illustrate layers of data but may become complex.
    - **Treemap**: Offers a hierarchical view but could be less intuitive for showing proportional relationships.
- **Chosen Chart & Detail**: A sunburst chart was selected for its ability to represent hierarchical relationships and proportions effectively. The chart categorizes the reasons for spills in both industrial and residential zones, with color gradients to differentiate between categories and subcategories. Design choices aimed to make the data easily digestible, with clear labels and segmentation to facilitate quick understanding of the predominant reasons for spills in each zone.

LIBRARIES

In this project, I strategically utilized a suite of Python libraries, each chosen for its specific functionality:

1. **Pandas**: Essential for data cleaning and aggregation, I employed Pandas to streamline the dataset of over 200k entries, focusing on organizing and grouping spill data since 1990 for different charts.
2. **Matplotlib**: This library was key for creating the boxplot and bubble chart. I leveraged its precision in plotting the trendline in the boxplot and accurately representing spill quantities in the bubble chart.
3. **Plotly**: I used Plotly for the interactive sunburst charts, allowing an exploratory view into the reasons for spills in different zones. Its dynamic capabilities brought an engaging element to the data, especially useful for web-based presentation.
4. **Folium**: For mapping spill locations, Folium's integration with Leaflet.js was invaluable. I created interactive maps to visualize spills in relation to geographic features, displaying Plotly Piechart onto the map.

5. **Streamlit**: For the web application, Streamlit was instrumental in transforming the data visualizations into an interactive, user-friendly format, making the findings accessible and engaging for the audience.

WORK SUMMARY

Task	Duration (in hours)	Month
Exploring other Project Ideas	6h	Oct-Nov 2023
Conceptualization and Data preparation	8h	Dez 2023
Boxplot visualisation	1h	Dez 2023
Initial Bubble chart	2h	Dez 2023
Gathering Feedback	3h	Dez 2023 – Jan 2024
Mapping Piecharts and Circles	10h	Jan 2024
Sunburst Chart creation and Data selection for visualisation	1h	Jan 2024
Updating Bubble chart	2h	Jan 2024
Prepare Website	2h	Jan 2024
Final Report	4h	Jan 2024

WORKS CITED & BIBLIOGRAPHY

Environmental Protection Authority Victoria. (2020, June 29). *Environmental Protection Authority Victoria*. Retrieved from About hazardous leaks and spills: https://www.epa.vic.gov.au/for-business/find-a-topic/about-hazardous-leaks

Etkin, D. (2001). Analysis of oil spill trends in the United States and worldwide. *International Oil Spill Conference*, (p. 11).

Hernandez, S. S. (2020, August 21). *Collision course: How a Japanese bulk carrier strayed from shipping lanes before disaster struck*. Retrieved from Reuters Graphics: https://www.reuters.com/graphics/MAURITIUS-ENVIRONMENT/SPILL/xlbvglnqrpq/index.html

ITOPF. (2021). *Oil Tanker Spill Statistic 2020.* ITOPF.

Llamas, M. (2023, April 26). *Chemical Spills*. Retrieved from ConsumerNotice: https://www.consumernotice.org/environmental/chemical-spills/

NY Open Data. (2024, January 5). Spill Incidents. New York State, USA: data.ny.gov.