SDS 3786 Lab 2: La Visualization des données.

Professor: Patrick Boily
Anthony Le, John Bute, Kelly Gao

Key Variables for Understanding the Datasets:

Pimento_Cases:

- **Time Series Variables (month, day, year):** These indicate the dataset logs daily activity (Monday-Friday) where employees record cases worked on throughout the day.
- **Death:** This column is crucial as it reflects the number of death-related cases, similar to other columns like Assistance Communications and Financial Assistance/Transfers.
- **Death Time:** Indicates the total time spent working on death-related cases during a specific day, similar to columns like Disaster Time and Financial Assistance Time.
- **GeoRegionNameE:** Highlights the operational span across regions, allowing insights into case distribution by region.
- **EmployeeID:** Confirms the logs represent work done by individual employees on specific days.

Pimento_Programs:

- **Time Series Variables (month, day, year):** These classify the amount of work done daily and link PIMENTO_PROGRAMS with PIMENTO_CASES.
- **MissionTitleE:** Provides context by identifying which projects employees are working on, such as embassy-related services.
- **EmployeeCode:** Helps clarify that the logs depict an individual employee's daily work on specific programs.
- **GeoRegionNameE:** Offers insights into geographical work distribution on particular dates.
- **Program_Mgmt:** Represents time spent on program management, with the highest non-zero data among similar columns, indicating a focus on management activities.

Atrocious visualizations of data and their issues

Pimento_Cases:

1.1 The follwoing bubble chart plots Employee IDs against Legal/Notary Time by month, with bubble size scaling by 5000 units. However, Employee ID is a unique categorical identifier, not a quantitative numerical value, and the months are not specified by year. The implied negative correlation and poor choice of image (selecting a bubble chart for the data storytelling) may mislead executives in performance reviews, suggesting employees with higher IDs work less. In reality, larger ID numbers may simply reflect newer employees or specific case categories, not workload or efficiency. The chart also violates Gestalt principles of similarity and continuity, using arbitrary colors and a non-sequential x-axis.

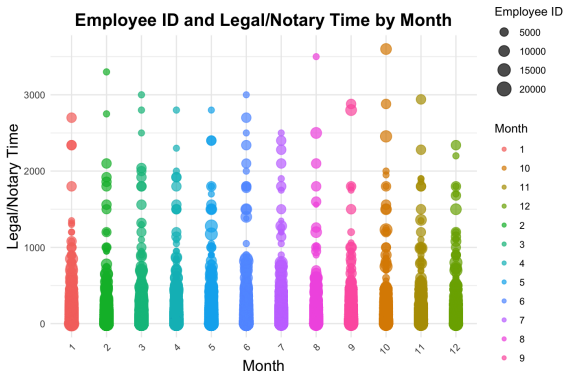


Figure 1.1 Poor visual 1

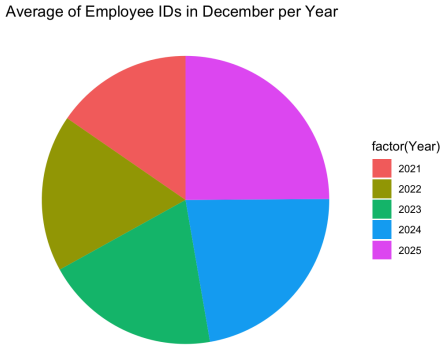


Figure 1.2 Poor visual 2

1.2 This pie chart shows the average of Employee ID numbers in December each year. Pie charts are meant to display part-to-whole relationships, but in this context it is an overused trope. The average Employee ID carries no meaningful numerical value, and the lack unit labels removes the ability to gauge any proportional magnitude or clear measurements. This chart would confuse managers when presented in a corporate setting- it fails in communicating a compelling visual story and fails to meets principles of analytical design by lacking context or relevant narratives.

Pimento_Programs:

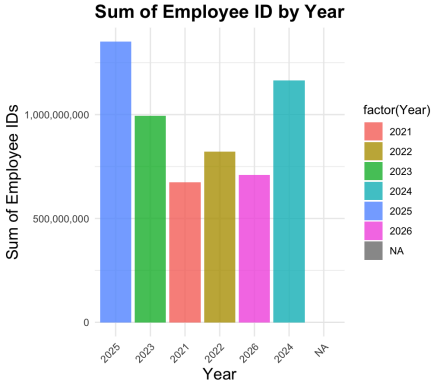


Figure 1.3 Poor visual 3

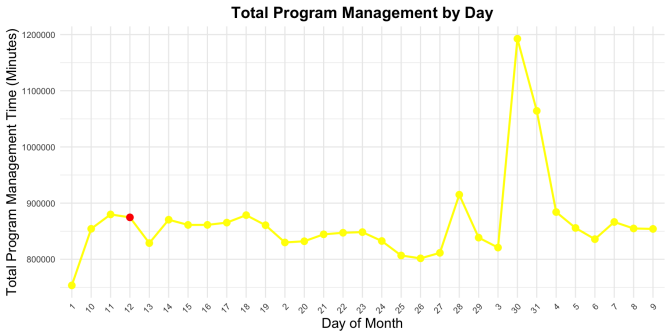


Figure 1.4 Poor visual 4

1.3 This bar chart maps the sum of Employee IDs each year. There lack of continuation among the bars and arbitrary colour use increases the cognitive load of visual processing. Moreover, the y-axis' variable of Employee IDs is aggregated into a sum, a meaningless metric as Employee IDs are unique identifiers and carry no summative significance. The choice of variables demonstrates poor analytical design by failing to consider storytelling elements.

1.4 This line graph plots total program management each day of the month. The use of a bright red dot on Day 12 draws focal attention to an otherwise insignificant data point; the low-contrast yellow line and background colour showcase poor visual duality. The lack of continuity on the x-axis leads to spikes or drops in program management that don't exist, and days are not distinguished by months or years as all collective months are lumped together. In a governmental setting when reassessing resource allocation, a public sector review may conclude certain days of the month having drastic drops in program management. This could cause misallocation of resources to false problem areas or enacting disciplinary measures against project managers.

Multivariable charts on PIMENTO_CASES

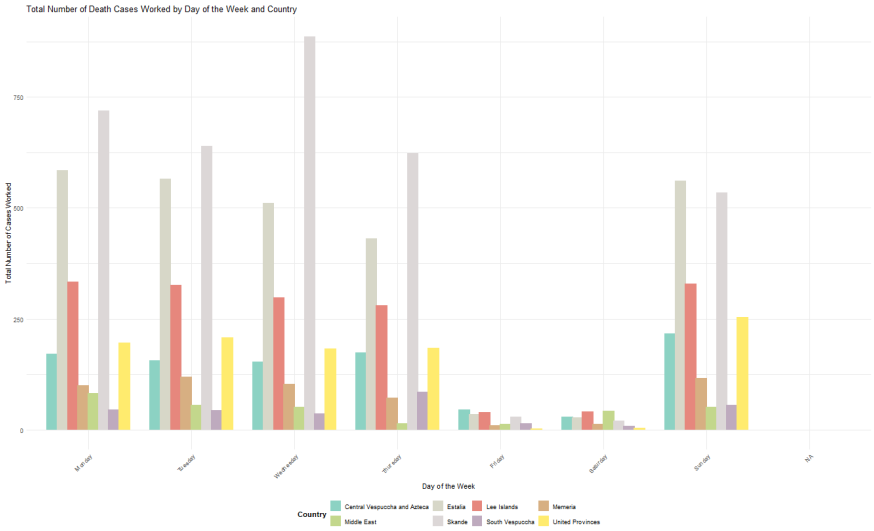


Figure 2.1 Multivariable visual for PIMENTO_cases

This multivariate visualization shows the total number of death cases handled by Canada's embassies across various regions, categorized by days of the week. The data highlights the workload for each region, along with how it fluctuates by day. To enhance readability and provide a broader overview, regions like Skande (West) and Skande (East) are combined under a single label, Skande.

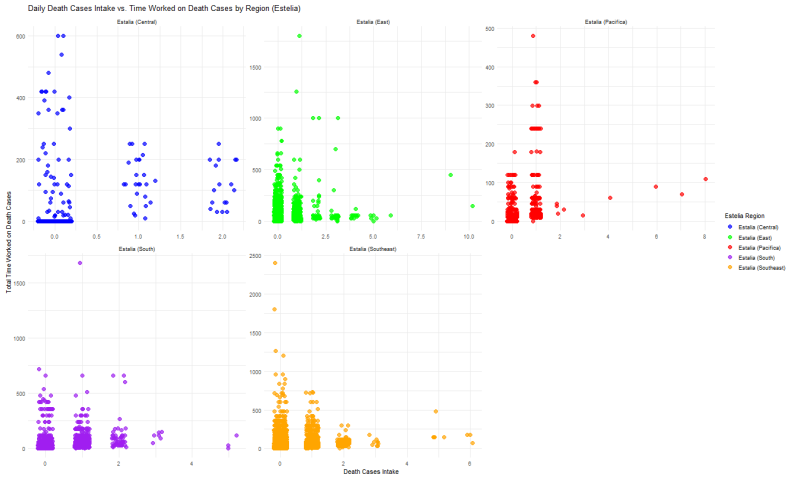


Figure 2.2 Multivariable visual for PIMENTO_cases

This graph visualizes the relationship between the time spent on death cases vs the amount of new death cases worked on during a day, spread out over the specific regions of Estelia.
Each region is represented with a separate visualization, where each dot represents a specific day within the region.

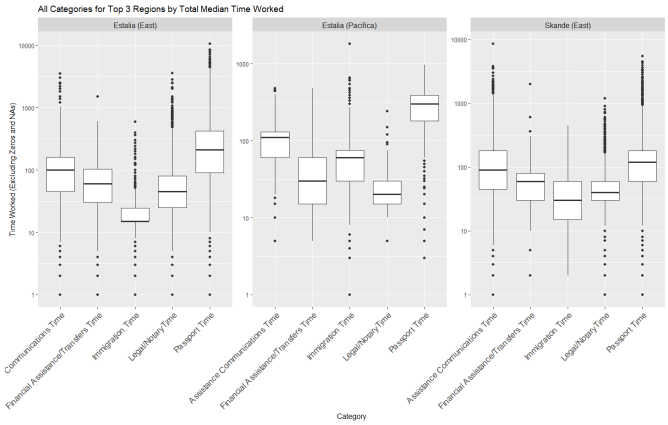


Figure 3.1 Multivariable visual for PIMENTO_cases

This image shows box plots comparing work time categories across three regions, focusing on daily time spent on specific case categories. The chart highlights the median (black line), outliers (dots), quartiles (25% and 75%), and data range (excluding outliers). The three regions with the highest median total work time were selected, and a log 10 transformation was applied to the y-axis for easier interpretation.

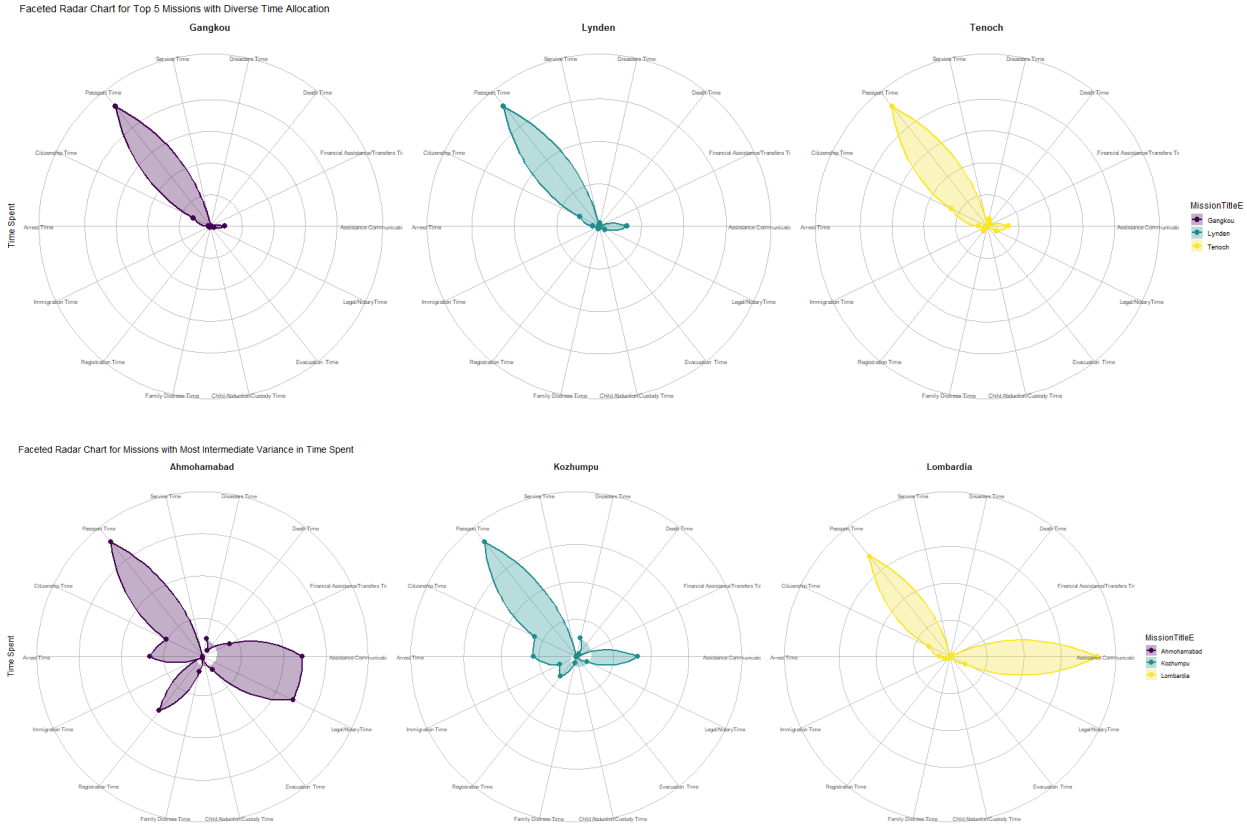


Figure 3.2 Multivariable visual for PIMENTO_cases

We created two sets of spider charts: one for the top 3 most active missions (missions spending the most time on cases) and one for the top 3 missions with the most intermediate variability (even distribution of time across case categories). This approach highlights which activities take the most time, notably passport cases, while also showcasing missions with a balanced focus across other categories.
We selected missions with variance closest to the median to ensure the radar charts reveal distinctions in time allocation across categories.

Multivariable charts on PIMENTO_Programs

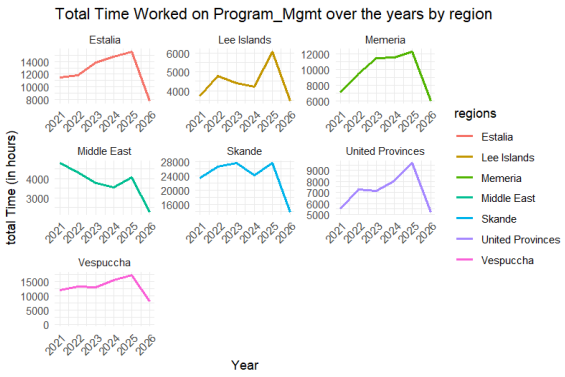


Figure 4.1 Multivariable visual for PIMENTO_Programs

This visualization depicts the total amount of logged time for Program_Mgmt in hours from all employees for each region, and shows the overall behaviour regarding the total amount of time they have worked. This approach to data visualization allows us to understand the correlation between each region and the amount of general work it took for the employees to fulfill their missions within these regions. We selected Program_Mgmt as the variable to measure against the total time spent as the variable was a more broad and more active variable than others.

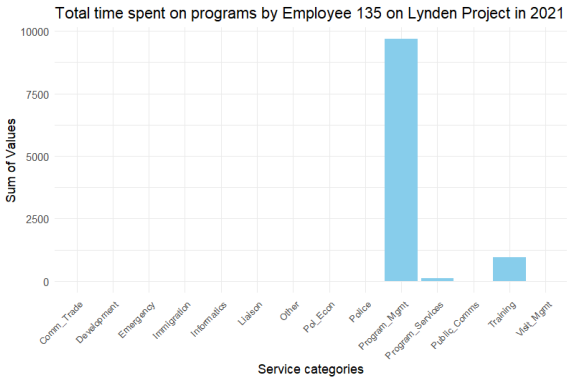
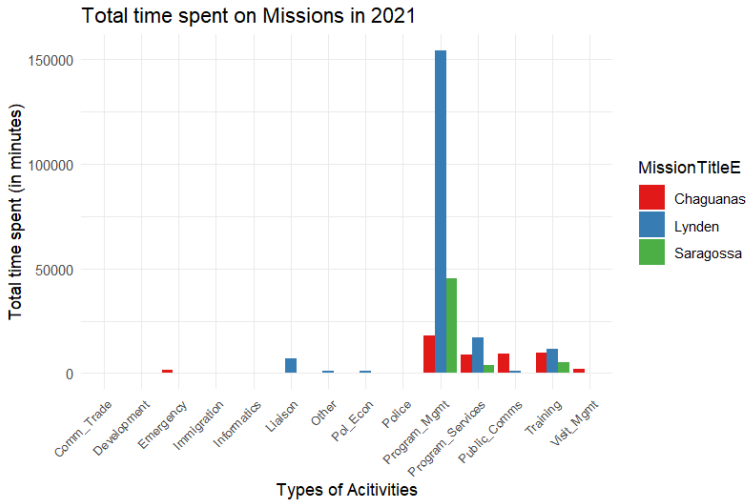


Figure 3.2 Multivariable visual for PIMENTO_Programs

This graph demonstrates the total amount of time spent by employee #135 on a variety of services for the Lynden Project in 2021. This visualization allows us to understand the employee's overall focus for their work, which would also allow us to understand their role and the role of the mission title overall. In retrospect, using the total activity time for all employees under the Lynden project would have been more effective to understand the project's goal, but due to time constraints, this was not possible.



4.1 Multivariable visual for PIMENTO_Programs

Similar to the previous graph, this visualization focuses on three different projects and the total amount of time spent on the kinds of work surrounding them. Using visualizations like this will allow us to understand the different aspects of the dataset and the different focuses that each mission title has on their respective regions, while also comparing them to other respective missions and their activities. Following a deeper dive, we could also have the same graph show the same stats for each region to determine a global understanding.

4.2 Multivariable visual for PIMENTO_Programs

This box plot diagram is meant to demonstrate the three most-worked employees in the “Stuvyessant” mission in 2021, depicting box plots for the three employees’ time spent (in hours) on Program_Mgmt. This box plot arrangement can be used to extrapolate different aspects about the mission, whether it was size, workflow, and mission goal. There, however, seems to be an issue with the data, as the box chart is depicting the hardest working employees logging in more than 100 hours in a single day, which is essentially impossible. We hope to discover more about this curiosity in our future labs.