

Validation and Refinement

Section 1. Domain-Level Validation (Who is the user?)

Step 1: Audience & Intended Action

The audience is new or incoming residents of the Los Angeles area who want to better understand the crime patterns of areas in the city that long-term residents may not be as interested in. Based on our problem and audience needs, the most important action we defined in PD2 is: "New or incoming LA residents will avoid hotspot areas and times with respect to crime in order to increase safety, make data-driven decisions on housing and commuting, and plan daily routines." This action is important because this specific group of LA residents is not fully versed on the crime patterns of the city, so a dashboard that portrays these analytics is crucial to reduce uncertainty and help them make informed decisions.

Step 2: Insights from Visualizations

- 77th street has the greatest number of reported crimes (15,050 crimes)
 - Most common crime code in this area was 624
- Central is the second largest number of reported crimes (14,752 crimes)
 - Most common crime code in this area was 624
- Lowest occurrence of crime is 5 am (7 cases on average) and the highest occurrence of crime is 12 pm (46 crimes on average)
- Crime rates fall from 12 am-5 am, pick up from 5-8 am, fluctuates from 8 am-11 am, peaks at 12 pm, then fluctuates the rest of the day.
- Firearms and knives are overwhelmingly used in violent crimes
 - Shotgun (90%), Semi-automatic handguns (85%), revolvers (75%), and knives (80%)
- Other weapons (screwdrivers, chemicals, chains, or household objects) commonly used to break into property
- From 2010-2018, crime dropped in 2013 (192,596 crimes) and peaked in 2017 (231,020 crimes)
 - Crime rates drastically increased from 2014-2017 and have been on decline since
- Fridays and Saturdays show slightly higher crime reports compared to other days
- Most common crime codes across the 21 LAPD stations were 510, 624, 440, and 310
- Top 10 crimes: Burglary, theft of identity, theft-grand (\$950.01 & Over), vandalism (\$399 or under), vandalism (felony \$400 & over all church vandalisms), theft plain (petty \$950 & under), burglary from vehicle, Battery (Simple assault), robbery, and intimate partner simple assault
 - Five categories: Burglary, theft, vandalism, assault, and robbery
- Burglary: Average victims aged 44 and 47

- Theft: Average victims aged 44, 42, and 39
- Vandalism: Average victims aged 42 and 41
- Assault: Average victims aged 32 and 24
- Robberies: Average victims aged 33

Step 3: Alignment Reflection

1. Do the insights from visualizations support the action defined earlier?

The insights provided align with the actions that we defined for the audience. I aim to inform newer residents of the areas with the highest frequency of crime and the time frames they should avoid these areas. The visualization of the crime hotspots gives a map of the LA with areas with the highest crime frequency, and their crime code are provided.

2. Are there any gaps that would make it difficult for the audience to act?

I did not identify any gaps in the data which would make it difficult for the audience to act. I created three visualizations that provide a clear and concise overview of the facets of crime in LA. The data provides the audience with the rates of crime across different dates, times, locations, and weapons along with who is more likely to face what type of crime.

3. What additional data or analysis might be needed before finalizing your dashboard design?

To better inform our audience and further influence their actions, I could add two pieces of data on the rate of crime based on gender and crime frequency during holidays. Accounting for gender when evaluating crime will allow for a deeper understanding of who is more likely to experience crimes besides just their age. The holidays are unique in the crime frequency committed during those time frames. Including this data could help people who are visiting LA for the Holidays or residents who aren't used to busier areas during these times.

Section 2: Abstraction-Level Validation (What is shown and why?)

Step 1: Restate Your Key Tasks and Data Attributes

Tasks:

- Identify areas with the highest frequency of crime (hotspots)
- Explore the temporal distribution of crimes to detect patterns
- Analyze severity by weapon type
- Compare crime rates across years to track trends

- Identify peak times of day and days of the week for various crime types
- Compare crime rates by type across various areas to determine if certain areas are prone to specific types of crime
- Profile average victim age by top crime codes

Attributes:

- Area names
- Crime count
- Date of crime occurrence
- Time of crime occurrence
- Weapon description
- Crime code
- Year (we derive this from DATE OCC)
- Crime type
- Average victim age

Step 2: Link to Audience Action and Insights

We selected these tasks to help new or incoming LA residents, who do not have much knowledge about safety and crime-related activities in the city, identify which areas and times are crime-dense and which types of crimes occur, to avoid risky situations. This supports our defined audience action: “New or incoming LA residents will avoid hotspot areas and times with respect to crime in order to increase safety, make data-driven decisions on housing and commuting, and plan daily routines.” Our exploration of spatial and temporal attributes reveals trends that guide users on which areas and times are the riskiest. Our abstraction can still be worked upon to explore other attributes that can help give new LA residents a fuller picture of the crime scene in the city.

Step 3: Identify Gaps in Data or Task Abstractions

In evaluating the crime trends over time, we failed to include the differences across different months of the year in addition to the crime rate that could be elevated due to holidays and special events. This can create a gap in the audience’s knowledge for specific dates/timeframes that could be high-risk.

Step 4: Propose Revisions or Extensions

Gap 1: No gender variable in the tasks

The absence of gender as a variable in the dashboard would create gaps in our audience’s ability to identify who is impacted by what crimes.

- New Task:
 - Compare gender distribution of people affected by different crimes.
- New Data attribute added
 - Gender

Gap 2: Lack of information related to holidays

With no knowledge on the holidays, the audience has no way to know whether they are going to be at a higher risk than usual due to special circumstances.

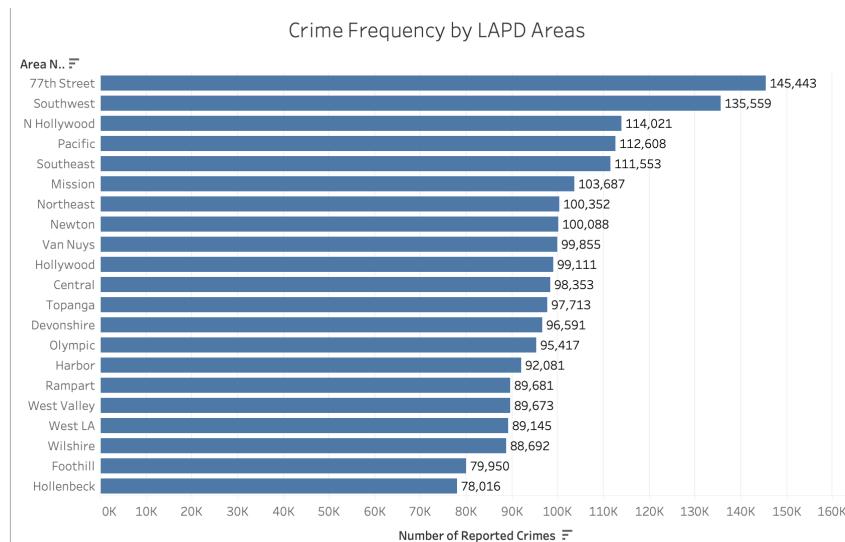
- New Task:
 - Compare the average frequency of crime throughout a year to the crime frequency during several major holidays.
- New data attribute needed:
 - Holidays

Section 3: Idiom-Level Validation (How is it shown?)

Step 1: Inventory of Visual Idioms Used (Table or list)

Task	Attributes Used	Visual Idiom
Identify Crime Hotspots	AREA NAME, Crime Count (Derived)	Bar Chart
Examine Temporal Crime Patterns	DATE OCC, TIME OCC, Crm Cd, Average Crime Count per hour (Derived)	Time Series
Analyze Crime Severity by Weapon Type	Weapon Desc, Crm Cd, Crime Count (Derived)	Stacked Bar Chart
Examine Crime Trends Over Time	DATE OCC, Year (Derived using DATE OCC), Crime Count per year (Derived using Year)	Time Series
Assess Crime Risk by Time of Day and Week	TIME OCC, Crime count (Derived)	Heatmap
Investigate Crime Hotspots by Crime Type	Area Name, Crm Cd Des, Crime count per area (Derived)	Map
Identify the Average Age of the Victim Affected Characterized by Crime	Crm Cd (Derive Top Ten frequent), Crm Cd Desc, Average Victim Age (Derive Vict Age and group by Crm Cd)	Bar chart

Step 2: Assess the Effectiveness of Each Idiom



Bar Chart:

Using a bar graph to identify crime hotspots was beneficial from an absolute judgment perspective. We have the bars labeled, making it easier to distinguish the areas with the most crime. Looking at the graph, we see that 77th Street and Southwest have the most crimes. Then, we notice Foothill and Hollenbeck as areas with the least crimes. Since this bar chart is in descending order, we can easily see the rankings of each area.

This graph may look a little overwhelming at first glance because all the bars are labels. Since our goal is to compare the LAPD areas crime, we might declutter this graph by only labeling the graphs that are like each other (less than 1,000 crime frequencies apart).

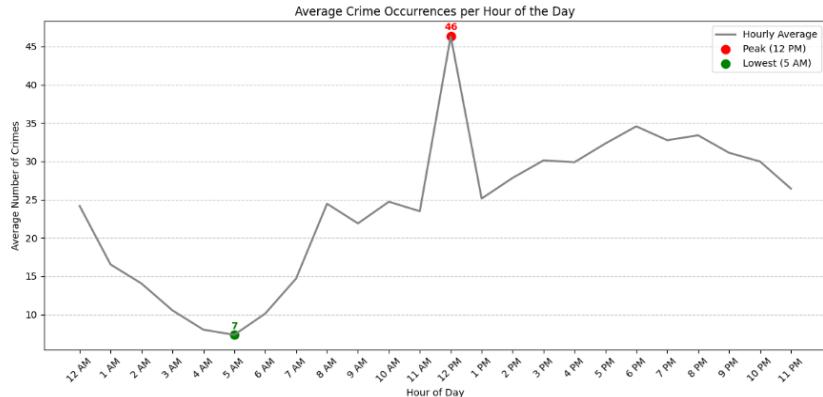
We can also use a color grid to categorize the crime frequency. Using sequential colors can help show severity in hotspots.

Light Blue: Crime Frequency 70,000-89,999

Medium Blue: Crime Frequency 90,000-109,999

Darker Blue: Crime Frequency 110,000-129,999

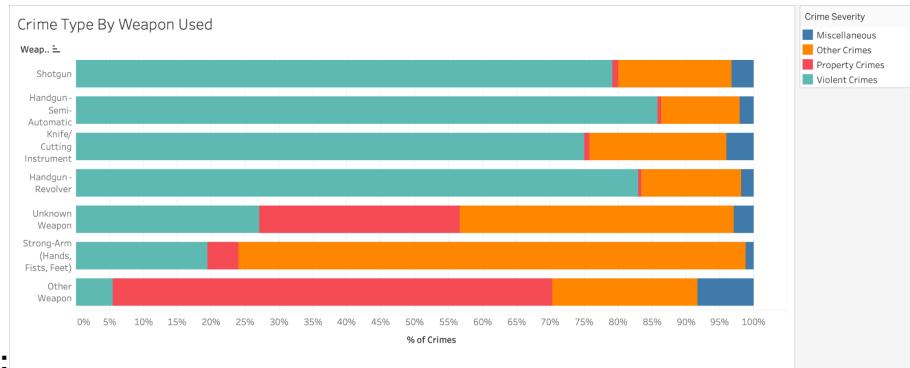
Dark Blue: Crime Frequency 130,000-149,999



Time Series:

Using a time series graph to identify temporal crime patterns throughout the day was very efficient. We looked at the average crime occurrence per hour of the day to help locate when crime peaks and when it calms down. We are very successful in incorporating this because we strategically labeled the peak time with a red and the lowest with a green dot. We also created redundancy in the graph, because we labeled what the average crime rate was at those specific times. The peak time for crime is 12 pm (46 crimes on average) and the lowest is at 5 am (7 crimes on average). Then, we also had the times listed in the color legend. This made the graph easy to interpret and helps direct the audience's eyes to specific times and crime rates.

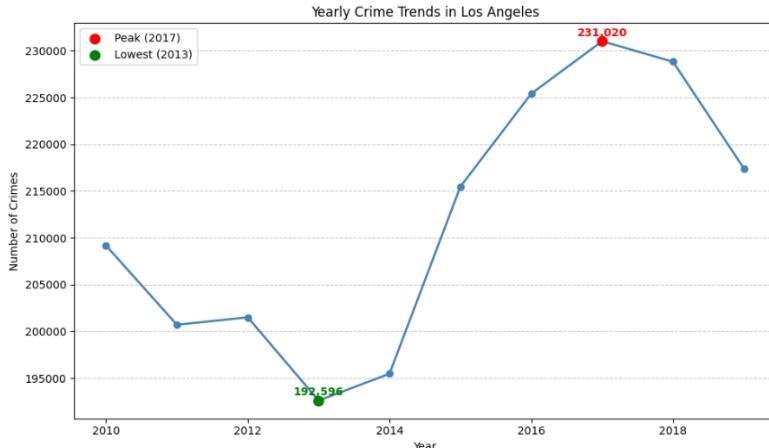
Through our use of color and other channel choices, our graph did not feel cluttered. It is appropriate for the audience to understand.



Stacked Bar Chart:

Using a stacked bar chart was very efficient for looking at weapon use because of its ability to do part-to-whole comparison. We can compare the different weapons used for different types of crimes. Through the colors, we can see the distribution of which weapons were most used per crime. We see shotguns/handgun-semi automatics, knives/cutting instruments, and handgun-revolvers used mostly for violent crimes. Therefore, we could see patterns within our graph. In general, the graph was easy to interpret and did not feel too cluttered.

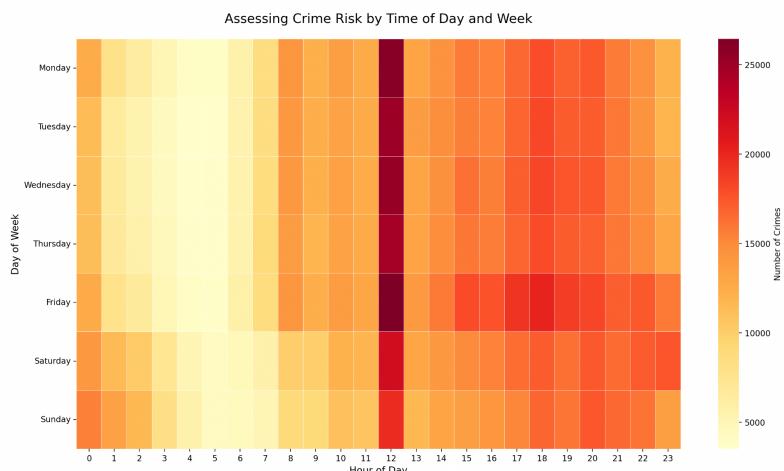
The only aspect our team questioned was how helpful it is for new residents to know which weapons are used for different crimes. This might be helpful for storeowners, law enforcement, or people to identify suspicious activity. However, in the future, we are going to remove this task.



Time Series:

Using a time series graph to examine crime trends over the years was very efficient. We looked at the number of crime occurrences through the years to identify which years crime peaked and fell. We are very successful in conveying these patterns to our audience because we strategically labeled the peak year in red and the lowest year in green. We also created redundancy in the graph, because we labeled the crime number for those years. The peak year for crime was 2017 (231,020 occurrences) and the lowest was 2013 (192,596). We also had the years listed in the color legend. This made the graph easy to interpret and helps direct the audience's eyes to specific times and number of crimes.

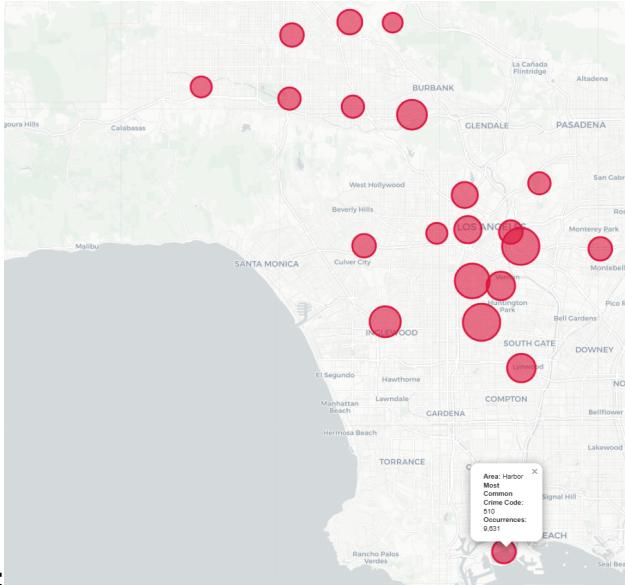
Through our use of color and other channel choices, our graph did not feel cluttered. It is appropriate for the audience to understand.



Heatmap:

Using a heatmap, we could successfully understand the crime risk based on time or day and week. In our earlier task, we identified which times crime peaked. This task is a type of extension because we are now looking to see if crime rates change based on the weekday. We see in the graph, that no matter what the day is, 12 pm is a peak time for crime. However, Fridays at 12 pm there are more crime occurrences than Sunday at 12 pm. Therefore, this visual is good for understanding the nuances of how the day of the week affects the number of crimes.

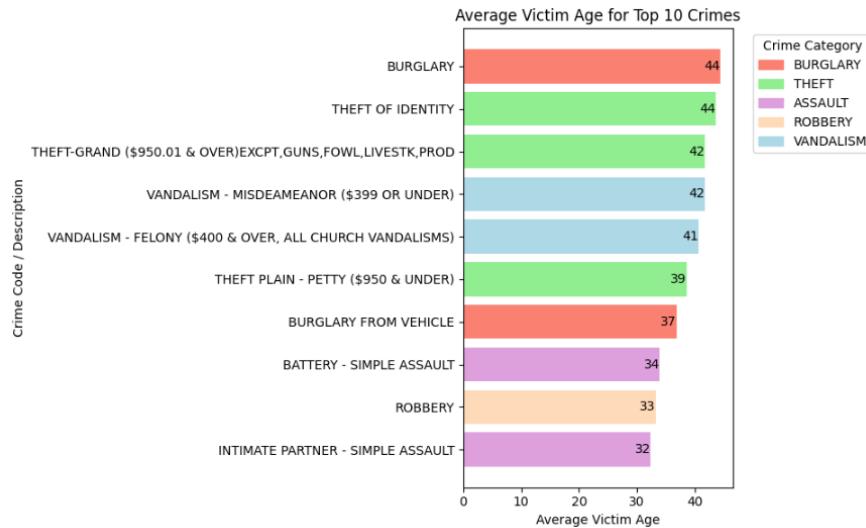
This heatmap is easy to interpret because the audience can look at the color hue to identify more crime risk. It conveys the pattern of what days and times crimes peak efficiently to our audience.



Map:

Using a map is especially helpful identifying what the crime hotspots are and the most popular crime type occurs. A map is important, because we can see if the LAPD areas with the most crime rates are next to each other or not. If you hover over the bubble, you can Area, Most Common Crime Code, and number of crime occurrences. Having this map makes it easy for the audience to find the hotspots based on the size of the bubble. They will also know what the most popular crime type is by hovering over an area.

The graph is not cluttered. The only improvements I would suggest is having the bubbles different colors based on sizes. I also think we should show the description of what the crime code means for the audience to understand.



Bar chart:

Using this bar chart is easy to understand the patterns behind which ages are most targeted. The use of colors for categorizing also helps an audience comprehend what types of crime are occurring. This is important because some of the top 10 crimes have the same crime listed but have a different level of severity. We also see redundancy in labeling the average victim's age, making it easier for comparison across categories. All of these aspects ensure that the visual is clear to interpret and patterns are easy to locate.

This visual is not cluttered. All of the visual channels that are added are there for a reason and help the audience understand.

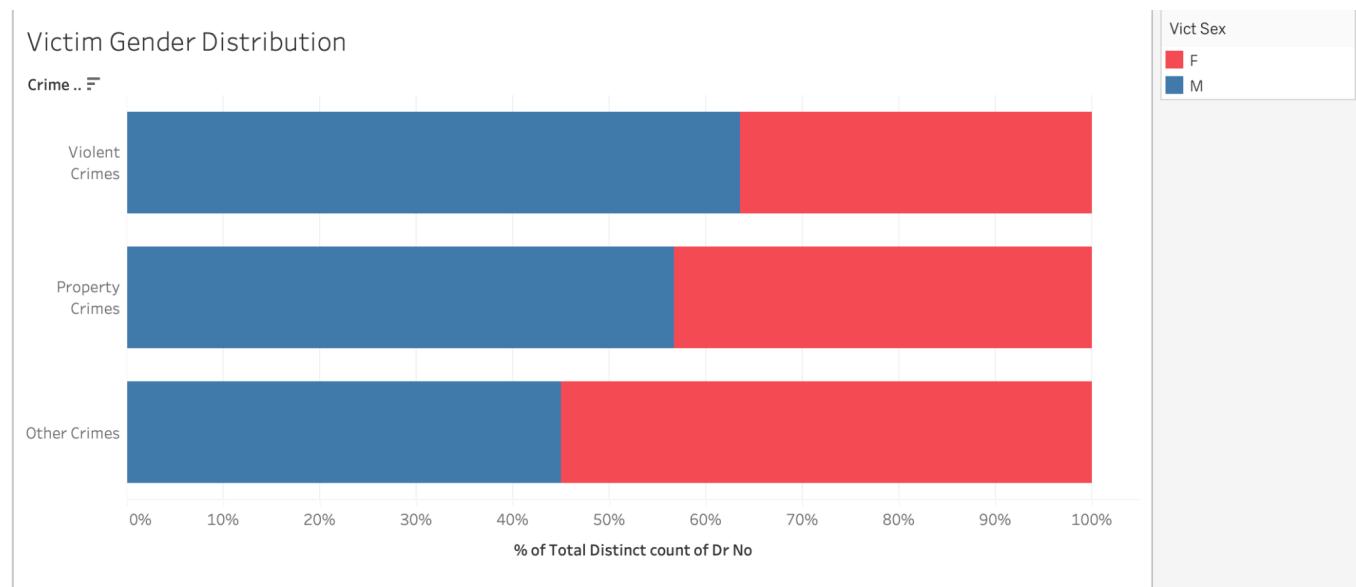
Step 3: Reflection and Recommendations for Improvement

Task	Visual Idiom	Status	Reasoning
Identify Crime Hotspots	Bar Chart	Revise Channels	<p>This visual is great for comparing the areas of crimes together. However, it is a little cluttered. We already have a map that shows us the crime hotspots, common crime codes, and crime occurrences. This visual is solely meant for general comparison in the number of crimes. Therefore, it may be helpful to not label all the numbers. Instead, we can use a color grid:</p> <p>Light Blue: Crime Frequency 70,000-89,999 Medium Blue: Crime Frequency 90,000-109,999 Darker Blue: Crime Frequency 110,000-129,999 Dark Blue: Crime Frequency 130,000-149,999</p>

Examine Temporal Crime Patterns	Time Series	Keep	New residents can easily look at the green and red dots and follow the trend line to see the average number of crimes throughout the day. This visual is easy to interpret and is applicable to new residents who want to know what times crimes most occur.
Analyze Crime Severity by Weapon Type	Stacked Bar Chart	Remove	In our analysis, we realize that identifying weapons and analyzing the severity of a crime will not help our audience. Our audience is not concerned about what weapons are being used to commit crimes. They want to know where the crime is located, what crimes are occurring, and how they will be most affected. Therefore, we will remove this task.
Examine Crime Trends Over Time	Time Series	Keep	New residents can easily look at the green and red dots and follow the trend line to see how the number of crimes has changed. This visual is easy to interpret and is applicable to new residents who want to know the background about where they are living.
Assess Crime Risk by Time of Day and Week	Heatmap	Keep	The heatmap shows new residents the various nuances in how crime risk changes based on the day. All the channels and visual choices are easy to interpret and show new residents' various trends.
Investigate Crime Hotspots by Crime Type	Map	Revise Channels	Currently, our map identifies the crime hotspots, the number of crime occurrences, and what crime code most occurs. However, the gap is that our audience does not know what the crime codes will be. Therefore, we will add a description of what the crime codes represent when we hover over any LAPD area.
Identify the Victim's Average Age by Crime	Bar chart	Keep	All the channels are being used efficiently to show new residents what ages are most targeted by specific crime.

Connecting Abstraction and Idiom: Refined Visualization Creation

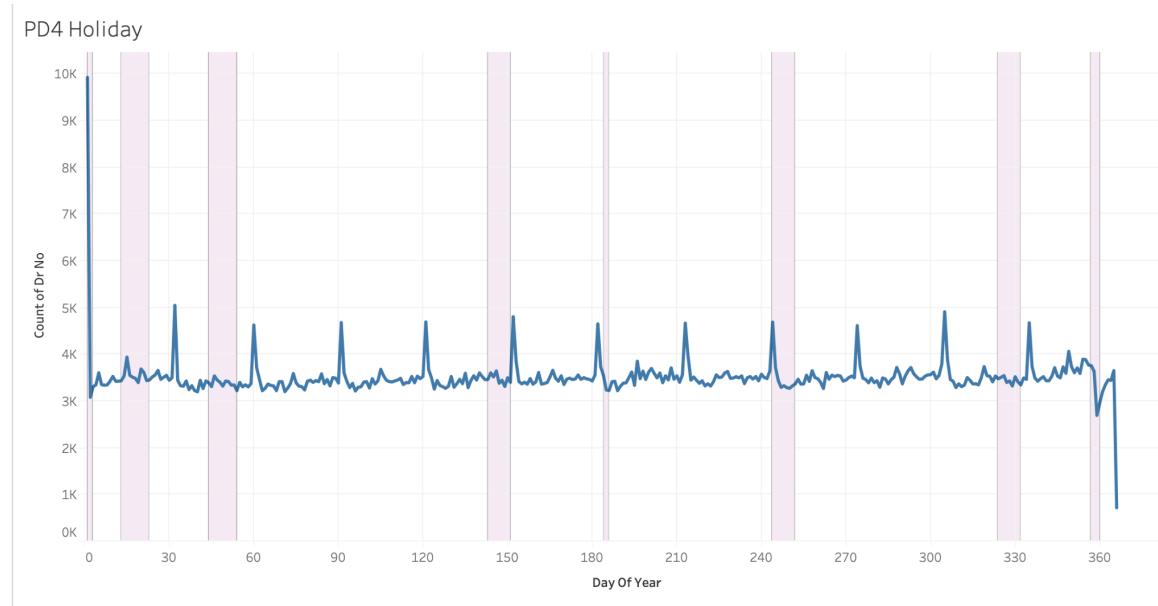
To address the first abstraction-level gap, the absence of gender-specific insights, we added a new task focused on comparing the gender distribution of victims across major crime categories (Violent, Property, and Other). Because this task requires part-to-whole comparisons within categories, we selected a horizontal stacked bar chart, which is well-suited for communicating proportional differences. Each bar represents a crime category and is divided into male and female victim proportions, ranging from 0% to 100% on the x-axis. This idiom makes it easy to compare gender patterns both within and across categories, supporting personalized safety interpretation for new residents who may want to understand how risk varies for people like themselves.



The visualization revealed clear gender differences across crime types. Violent crimes show roughly a 63% male to 37% female split, while property crimes are somewhat more balanced (57% male, 43% female). The “Other Crimes” category reverses the pattern, with females comprising approximately 55% of victims. This reversal suggests that specific offenses within this category disproportionately affect women and may warrant further exploration. The idiom effectively enabled detection of these proportional differences, directly supporting our audience’s ability to recognize which populations are more vulnerable in specific crime contexts.

To address the second abstraction-level gap, the lack of information about crime behavior during major holidays, we created a new task comparing the average daily crime count across the year with crime activity during major holiday periods (New Year's, MLK Day, Presidents' Day, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas). We used a time-series line chart, which is the most

effective idiom for revealing temporal trends, seasonality, and deviations from baseline behavior. Shaded reference bands mark the holiday periods, allowing users to quickly identify spikes or dips relative to the yearly average.



The resulting visualization revealed several notable temporal patterns. Daily crime counts show regular weekly fluctuations and mild increases at the beginning of each month. When examining the holiday bands, crime does not uniformly increase during major holidays. New Year's displays a sharp spike of over 10,000 incidents across the aggregated 10-year period, whereas the Christmas period shows a substantial drop to under 3,000 incidents. Memorial Day, Independence Day, and Labor Day show slight increases above the baseline average of approximately 3,500 crimes per day. By making holiday periods visually distinct, the line chart clearly communicates these deviations, helping new residents anticipate when crime risk is unusually high or low due to holiday-specific circumstances.