Crime and Weather

Jonathan Goins
University of Colorado
Minneapolis, MN
jogo8350@colorado.edu

Brady Murphy
University of Colorado
Gunnison, CO
brmu6084@colorado.edu

Lauren Paredes
University of Colorado
Stillwater, OK
lapa5112@colorado.edu

Dylan Ginsberg
University of Colorado
Denver, CO
dygi8890@colorado.edu

Abstract Summary

In this project, we aim to answer four questions 1. Do extreme weather days see more of a specific type of crime? 2. Does each season have a type of crime that occurs more frequently? 3. How do different cities in the U.S. compare to each other (i.e., different climates)? 4. How do cities across the U.S. compare based on crime types classified by person, property, or society? To do this, we classified crime and weather data sets for four cities across the US, Denver, CO; Boston, Massachusetts; Los Angeles, California; and Seattle, Washington. Focusing on the weather, we decided to make temperature groups based on the max temp for each day which were hot (> 80) ℉, mild (50-80)℉, and finally cold (< 50)℉. Then the crime data sets were classified into three categories, property, person, and social crimes.

Over the semester, we went through five stages: prep, exploratory analysis, merging data, calculation, and interpretation. In our prep stage, we ensured each data set was similar and had enough data entries to be ready for other stages. In our exploratory analysis, we made plots to visualize our data to make educated guesses about potential findings. Then we had to merge our crime datasets with the weather datasets. This merging section was essential to our process, as this combined dataset would be what all the calculations are made of. Finally, we did our calculation phase, so we individually conducted general distribution and Bayesian classifications on each city. Then in our interpretation phase, we got together as a team and compared all four cities' calculations.

Introduction

The main goal of this project is to find the correlation between Crime and Weather and to predict what type of crime most often occurs given a temperature range. We use public crime datasets from Denver, Seattle, Boston, and Los Angeles in the United States from 2017-2022. Our findings will be helpful for local law enforcement and neighborhood watch groups. Additionally, these findings will generate public awareness about possible crime and allow people to plan. These findings will be found via Bayesian classification on a city level. Then all four cities will be compared to one another to get a broader view of our data. Some of the challenges we have encountered include: finding a way to merge crime and weather data. This was crucial for our project as this merged data set would be what we do our calculations on.

Another problem we encountered was encompassing as many crimes as possible into the three categories (Property, Person, and Society) with minimal incorrect classifications and missing data. The next major problem was finding temperature ranges that worked for all cities with a good distribution of crimes. Finally, the team's final challenge was creating and interpreting our results.

Related Work

There are several other projects closely related to our topic. By looking at these examples, we can better understand what others have done and how their results compare to ours. Related work also helps us understand how the data should be compared and how we can formulate trends based on this data. Most related work similar to our project looks at temperature vs. violent crimes. This is the easiest way to narrow down the data and get a good feel for the weather on a given day. Violent crimes were also seen as the most important when gathering what crimes to use. Our project categorized crimes into three categories instead of just violence. Some of these projects are older, and some are more modern; either way, these projects gave a great foundation to this analysis.

Example Projects:

- Mining the Relationship between Crimes, Weather, and Tweets
    - This project examines whether climate change can increase violent crimes by comparing crime rates at different temperatures from 1996 to 2013.
- Crime and Weather in South Africa
    - This project looks at the association between weather and crime in a township setting in South Africa from 2006 to 2016.
- Crime and Weather in Cleveland
    - This analysis examines the correlation between weather and violent crimes in Cleveland from 1999 to 2004.
- The troubling ways a heatwave can warp your mind
    - This BBC article looks at the increase in crimes due to temperature across the globe (comprehensive)

Hypotheses

For extreme weather we will have different types of extreme weather, and that is hard to say what we exactly suspect. However, we hypothesize that, at the very least, when the temperature is high, petty crime will be more prevalent. We expect to see more drug and narcotic offenses when the temperature is low.

Additionally, we will find that during the summer, car theft will be more prevalent; in the winter: robberies; in the spring: larceny and theft; and in the fall: drug and narcotic offenses.

Finally, comparing different locations in the United States, we theorize that places with more temperate and consistent climates will see more crime than those with more drastic seasons.

Preparing for our project, we made a lean spec to document our rules and classifications. Within our lean spec, we made a list of columns every city should have to maintain fluidity across our data sets. The columns included are date, time, temperature, and type of crime. We also classified our temperatures so that cold is less than 50℉, Mild is 50-80℉, and Hot > 80℉.

We have also broken up crimes into person, property, or society types. Person crimes involve/happen to another person (assault, murder, drug/alcohol). Property is classified as crimes that deal with one's property (robbery, arson, larceny). Society crimes include multiple persons or things that happen publicly (white-collar crimes, public disorder). For each city, these will be the guidelines for classifications. Additionally, there will be an error that accounts for the crimes that occur a few times that get categorized as others.

Pre-Processing

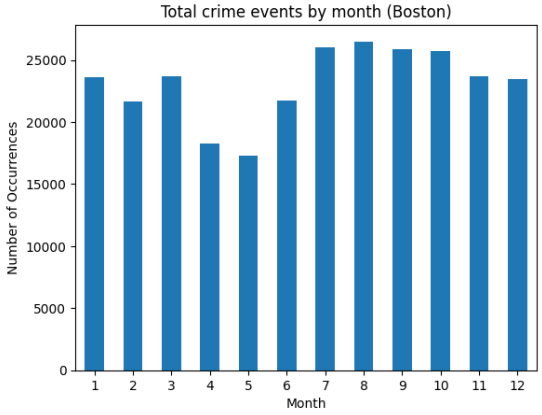
During our pre-processing, our goal was to demonstrate our prep as well as to be able to visualize our data. We first gathered how often a type of crime occurred for our crime datasets. We did this because we wanted a feel for what to expect as we continued.

Our next step was to create graphs to show crime counts over the multiple years contained within our datasets, then to break that further down, we aggregated crime counts per month.

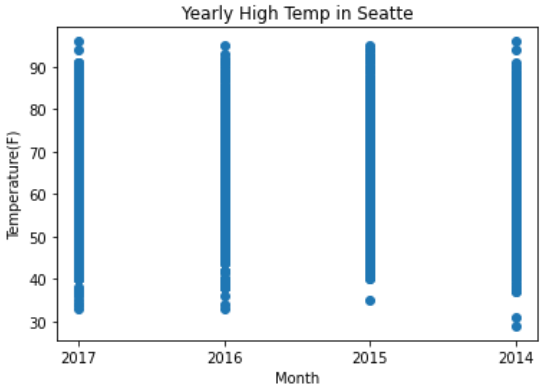
For our weather datasets, we had the same goals. We first saw the weather vs. time and then broke that down into the average monthly temperature, similar to what we did for our crime sets.

Additionally, we plotted precipitation for each of our cities to see if that could be something we consider in our next steps, as some of the cities we are considering have much higher precipitation than others.

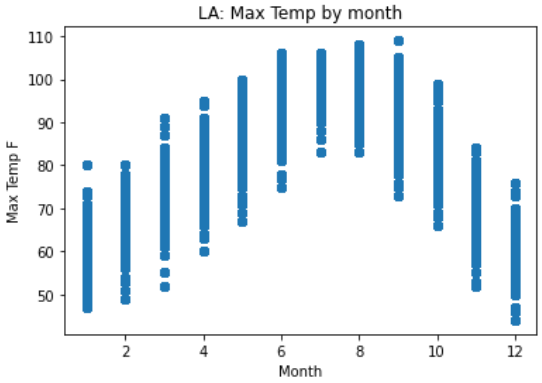
Denver crime counts per month



Boston crime counts per month



Seattle Average Temperature (℉) by month



Los Angeles Average Temperature (℉) range by year.

Finally, we found the years that each city had overlap in weather and crime: Seattle ~8 full years, Los Angeles ~9 full years, Boston ~2 full years, and Denver ~3 full years.

Methodology

In this project, we will strive to answer four main questions:

1. Do extreme weather days see more of a specific type of crime?
2. Does each season have a type of crime that occurs more frequently?
3. How do different cities in the U.S. compare to each other (i.e., different climates)?
4. How do cities across the U.S. compare based on crime types classified by person, property, or society?

To answer these, we will split the work process into several tasks. This was done by giving each member a dedicated city to look at and mine those data sets accordingly. Each city has a jupyter notebook depicting graphs and data frames to help visualize data and lay out the process we went through. These notebooks and visuals were shared over GitHub and google docs. Our primary form of communication was through the use of Discord.

The first main task, data preparation, will include joining the crime datasets with the weather datasets. Specifically, joining weather data for the given date to the date of the crime occurrences. This was hard to accomplish as not all data sets had a “date” column. This led to our next task;

constructing date columns. Once each data set had a date column that looked identical, we could merge crimes to weather days for the given city.

While mining for correlations, we specifically focused on our questions having different constraints on the data for each. The constraints will be 1. temperature groups as defined above (hot, mild, cold), 2. grouping crime types on person, property, and society, and 3. based on location in the United States. We decided to use “Max Temp” to classify our hot, mild, and cold days as we felt it showed a more accurate representation of the weather on a given day. In order to classify types of crimes, we used varying methods across cities. This led to missing crimes in the overall classification, but we agreed to keep missing percentages close in value and at a reasonable loss.

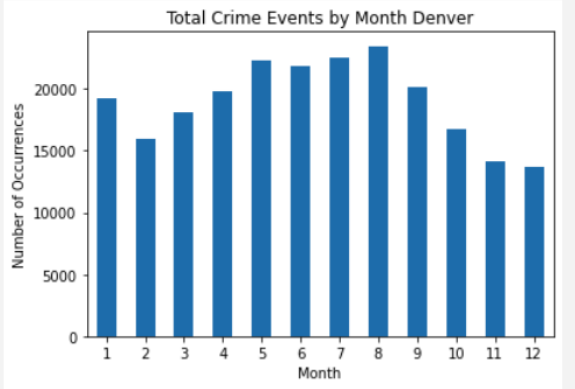
Once we had our crimes laid out by type and weather, we could start looking at relationships in the data. This was done using Bayesian Classification. Specifically looking at the three different days and the three different types of each day. Example: Given it is a hot day, how likely will the crime be property, person, or society?

The validation included mathematically proving that our mined correlations have enough of a confidence interval that the results can be considered significant. This also included comparing results amongst the cities to determine what guidelines fit all the cities best. This included things such as: finding high-temperature limits to classify days, how to classify crimes, missing crime percentages, and distribution of data for all cities.

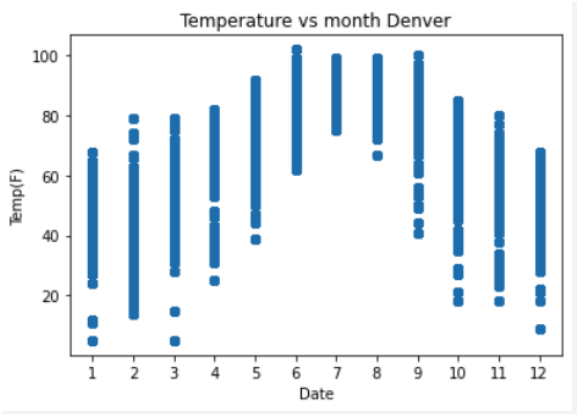
Our final task will be presenting our results. We want to include all four cities and their findings to determine our conclusions accurately. We can create big data from more minor data by comparing all four cities. This will be done in this final report and our presentation depicting our results.

Evaluation.

For the evaluation section of this project, there were three main elements: crime vs. weather over time, crime type distribution, and Bayesian classification of data. This is seen below with an example from each section. First, we see the total crime occurrences in each month over the several years included in the Denver data.



The following plot shows the max temperatures throughout the months also included in the Denver data.



The following two tables deal with the calculation aspect of the project. First, we look at crime distribution. This was an essential aspect between cities to ensure we all had an excellent data spread. It looks at the percentage of crimes across each category and day. The percent missing is the number of crimes for which we could not find a category. Some notebooks include all crimes, and some do not base on how the data was formed between police departments.

LA Crime Type Distribution	Hot	Mild	Cold	Total
Property	30.26%	31.97%	0.56%	62.786%
People	12.14%	11.65%	0.2%	23.984%
Society	3.06%	4.67%	0.05%	6.098%
Total	45.46	48.29%	0.81%	~92.868%

The following table is an example of each city's Bayesian calculations. It depicts the likelihood of each crime type on a given day.

Bayesian (High Temp (F)):

Boston	Hot	Mild	Cold
Property	50.54	49.31	47.73
People	37.33	38.37	39.66
Society	12.13	12.32	12.61
Total	100%	100%	100%

This project and its efficacy can be compared to the results found in similar studies of different cities, as mentioned above, such as the 2022 report published in the International Journal of Biometeorology on the associations between crime and weather in South Africa. As the report on South Africa has been thoroughly researched and edited, it will provide a stable backing to test conclusions and appropriate data modeling against this project.

The goal of compiling all this data is to be easily understood and interpretable by citizens and public servants local to the various cities. As such, this project's most significant evaluating factor is for its conclusions to be absorbed by the aforementioned groups and easily applied to help them be better prepared for future spikes or dips in crime rates following the weather patterns.

Milestones

This project had the following timeline:

1. Checkpoint Report - November 17th
2. Compare Cities - November 20th
    a. The remainder of the project focuses on finding relationships between cities, such as temperature ranges, crime rates, and crime types. By doing these for each city, we can determine when crime spikes are seen and if it relates to when average weather temperatures spike.
3. Validation - November 23rd
    a. We want to look at the types of crime that occur most often in specific temperature ranges. By doing so, we can correlate the types of crimes

throughout the year to the daily average temperature of said time of year.
    b. This will give insight as to how weather affects crime throughout the year.
4. Conclusions - November 30th
5. Final Report - December 6th

Discussion

Throughout this project, we all had to manipulate and clean our data differently because none of our data sets were identical.

Some things that worked well for us included splitting the workload, using Jupyter notebooks, using a clean-up tracker, and using Discord as our way of communicating with each other.

One thing that we had trouble with early on was a lack of communication and initiative.

In the future, look at classifications more precisely than just max temp but include perception of wind, temp, and rain. Additionally, looking at even more specific locations other than generalizing crime locations.

Finally, to find more accurate results. Study a broader range of years and consider events that may affect crime rates (COVID-19, BLM). Categorizing cultural and economic demographics would also relay more findings related to weather and crime.

Conclusion

Over all four cities, we saw Distributions: crime: Property crime on a mild day most common LA, Denver; Boston, and Seattle have similar distributions

The overall spread of crimes showed that property crimes are most common regardless of the weather. Also, more specifically, property crimes on a mild day are the most common out of all the three joint probabilities between crime type and temperature. Society crimes were the least common, however, except for Denver. All the cities showed that mild days were the most common temperature throughout the year and contained the most crimes. Mild days containing the most crimes allow us to conclude that late spring and fall throughout the year will have more crimes. L.A. and Denver have similar crime and weather distributions, while Seattle and Boston have similar distributions.

For our overall findings, we used two average value tables. The first showing the overall crime distribution and how much crime was encapsulated in the study. The second table shows us the average Bayesian. We included this table to help represent the seasons and given probabilities. These two tables were essential in determining our final findings.

Looking at the probability distributions outlined by the Bayes table, the probability of a society crime given any temperature is least likely to happen while the property is the most common. Property crimes are by far the most frequent across all weather days. Given that we have the most mild days, it is worth noting that both hot and cold days say more property crime than mild days. This was also the case with crimes against people.

It is important to note that the Bayes tables are simple conditional probabilities on weather and crime type, i.e., probability of society-type crime given hot weather - P(Scoiety|Hot). They are not complete Bayesian classifications on multiple features.

From our findings, we can conclude that weather does have an impact on crime. We see specific types of crimes more frequently given a different temperature day, showing some correlation to weather.

References

AnkurJain. "Crimes in Boston." *Kaggle*, 4 Oct. 2018, https://www.kaggle.com/datasets/ankkur13/boston-crime-data.

Butke, Paul, and Scott C. Sheridan. "An Analysis of the Relationship between Weather and Aggressive Crime in Cleveland, Ohio." *AMETSOC*, American Meteorological Society, April 1st. 2010, https://journals.ametsoc.org/view/journals/wcas/2/2/2010wcas1043_1.xml.

Department, Los Angeles Police. "Crime Data from 2010 to 2019: Los Angeles - Open Data Portal." *Data.lacity.org*, November 30th 2022, https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z.

"EBSCO Discovery Service: EBSCO." *EBSCO Information Services, Inc. | Www.ebsco.com*, https://www.ebsco.com/products/ebsco-discovery-service.

"How to Manage Pests." *California Weather Data: Set Dates and Variables--UC IPM*, http://ipm.ucanr.edu/calludt.cgi/PCSTATIONDATA?STN=EASTON-01.P1440&SCALE=Daily.

Mendes, Bruna. "Analysis of NYC Reported Crime Data Using Pandas." *Medium*, Towards Data Science, February 27th 2021, https://towardsdatascience.com/analysis-of-nyc-reported-crime-data-using-pandas-821753cd7e22.

Mooney, Paul. "Denver Crime Data." *Kaggle*, October 5th 2022, https://www.kaggle.com/datasets/paultimothymooney/denver-crime-data.

Mooney, Paul. "Denver Crime Data." *Kaggle*, October 5th 2022, https://www.kaggle.com/datasets/paultimothymooney/denver-crime-data.

Peng, Jennifer. "Boston Weather Data Jan 2013 - Apr 2018." *Kaggle*, April 10th 2018, https://www.kaggle.com/datasets/jqpeng/boston-weather-data-jan-2013-apr-2018.

Rl. "How to Pd.merge() Two Data-Frames on a Common Date Column." *Medium*, Analytics Vidhya, November 3rd 2020, https://medium.com/analytics-vidhya/how-to-pd-merge-two-data-frames-on-a-common-date-column-e7808d7ccaee.

Spd. "SPD Crime Data: 2008-Present: City of Seattle Open Data Portal." *Seattle Open Data*, December 4th 2022, https://data.seattle.gov/Public-Safety/SPD-Crime-Data-2008-Present/tazs-3rd5.

Swarr, Kaden. “Denver Hourly Weather.” *Kaggle*, March 18th. 2021, https://www.kaggle.com/datasets/kadenswarr/hourly-weather-data-in-denver.

Tatman, Rachael. “Did It Rain in Seattle? (1948-2017).” *Kaggle*, December 20th 2017, https://www.kaggle.com/datasets/rtatman/did-it-rain-in-seattle-19482017.

Tiihonen, Jari, et al. “The Association of Ambient Temperature and Violent Crime.” *Scientific Reports*, U.S. National Library of Medicine, July 28th 2017, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5533778/.

“The Troubling Ways a Heatwave Can Warp Your Mind.” *BBC Future*, BBC, https://www.bbc.com/future/article/20200817-the-sinister-ways-heatwaves-warp-the-mind.

Appendix

CU Boulder Honor pledge: “On my honor, as a University of Colorado Boulder student I have neither given nor received unauthorized assistance”

Individual Contributions: Overall, teamwork was evenly split up. Each member had a city and was responsible for their data; Lauren: Denver, Brady: Boston, Jonathan: Los Angeles, and Dylan: Seattle.