

Crime Dataset

Shelley Kim, Kristy Wang

Group 2

99-519: Discovering the Data Universe

Summary of dataset

- Each row is one recorded incident
- Columns used included
 - Time
 - Location
 - Neighborhood
 - Offenses

	INCIDENTTIME	INCIDENTNEIGHBORHOOD	OFFENSES	X	Y
1	2020-01-02 14:43:00	East Allegheny	2701 Simple Assault. / 3304 Criminal Mischief. / 3921 Theft ...	-80.00057	40.45350
2	2020-01-02 15:15:00	Homewood South	2705 Reckless Endangering Another Person. / 3304 Crimina...	-79.88781	40.45039
3	2020-01-02 21:00:00	Northview Heights	2702 Aggravated Assault. / 2705 Reckless Endangering An...	-80.00181	40.47805
4	2020-01-03 13:50:00	Beechview	2701 Simple Assault.	0.00000	0.00000
5	2020-01-05 15:18:00	Perry North	903 Criminal Conspiracy. / 2701 Simple Assault. / 2706 Terro...	-80.01111	40.47835
6	2020-01-06 12:41:00	East Liberty	2702 Aggravated Assault.	0.00000	0.00000
7	2020-01-02 17:23:00	Central North Side	1301 Registration and Certificate of Title Required. / 13(a)(3...	-79.98571	40.46940
8	2020-01-02 17:51:00	Homewood South	3802(a)(1) DUI - General Impairment / 9998 Overdose Other	-79.89471	40.46147
9	2020-01-02 16:00:00	Spring Hill-City View	3502 Burglary.	-79.99204	40.46536
10	2020-01-02 18:00:00	Allentown	9124 Drug Investigation	-79.99371	40.42180

Summary of dataset

- Arrests
- Each row is one recorded incident
- Columns used included
 - Time
 - Location
 - Neighborhood
 - Offenses
 - Gender, Race

GENDER	RACE	ARRESTTIME	ARRESTLOCATION	OFFENSES	INCIDENTLOCATION	INCIDENTNEIGH
F	B	2016-08-24T12:20:00	4700 Block Centre AV Pittsburgh, PA 15213	3929 Retail Theft.	4700 Block Centre AV Pittsburgh, PA 15213	Bloomfield
M	W	2016-08-03T14:55:00	4200 Block Steubenville PKE Pittsburgh, PA 15205	13(a)(16) Possession of Controlled Substance	4200 Block Steubenville PKE Pittsburgh, PA 15205	Outside City
F	B	2016-08-03T16:45:00	900 Block Freeport RD Fox Chapel, PA 15238	3929 Retail Theft.	900 Block Freeport RD Fox Chapel, PA 15238	Westwood
F	W	2016-08-05T02:36:00	Foreland ST & Cedar AV Pittsburgh, PA 15212	5503 Disorderly Conduct. / 5505 Public Drunkenness	Foreland ST & Cedar AV Pittsburgh, PA 15212	East Allegheny
M	B	2016-08-06T02:00:00	900 Block Woodlow ST Pittsburgh, PA 15205	2702 Aggravated Assault. / 2705 Recklessly Endangering Another Person. / 3925	900 Block Woodlow ST Pittsburgh, PA 15205	Crafton Heights

Additional Datasets

- Other geographic data
 - WPRDC GeoJSON files
- Categorizing Offenses
 - Web-scraping from PA Title 18

```
{  
  "type": "FeatureCollection",  
  "name": "Police_Zones",  
  "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },  
  "features": [  
    { "type": "Feature", "properties": { "FID": 1, "zone": 4,  
      "SHAPE_Length": 0.27885878957551241, "SHAPE_Area":  
      0.00096934374570333763 }, "geometry": { "type":  
      "MultiPolygon", "coordinates": [ [ [ [ [ -79.932848349999972,  
        40.396706622000067 ], [ -79.932751368999959,  
        40.396089946000075 ], [ -79.932583233999935,  
        40.394464873000061 ], [ -79.932309994999969,  
        40.391675663000058 ], [ -79.931152581999982,  
        40.390872208000076 ], [ -79.929574801999934,  
        40.389784533000068 ], [ -79.923994641999968,  
        40.394460655000046 ], [ -79.91815418799996,  
        40.399358707000033 ], [ -79.916210972999977,  
        40.398492722000071 ], [ -79.916161615999954,  
        40.398765679000064 ], [ -79.916074594999941,  
        40.398705007000046 ], [ -79.913244129999953,      ] ] ] } ] } ] }
```

GeoJSON file example

**ARTICLE B. OFFENSES INVOLVING DANGER
TO THE PERSON**

Chapter 23. General Provisions

§ 2301. Definitions.

Chapter 25. Criminal Homicide

§ 2501. Criminal homicide.

§ 2502. Murder.

§ 2503. Voluntary manslaughter.

§ 2504. Involuntary manslaughter.

§ 2505. Causing or aiding suicide.

§ 2506. Drug delivery resulting in death.

§ 2507. Criminal homicide of law enforcement officer.

Chapter 26. Crimes Against Unborn Child

§ 2601. Short title of chapter.

§ 2602. Definitions.

§ 2603. Criminal homicide of unborn child.

§ 2604. Murder of unborn child.

§ 2605. Voluntary manslaughter of unborn child.

§ 2606. Aggravated assault of unborn child.

§ 2607. Culpability.

§ 2608. Nonliability and defenses.

§ 2609. Construction.

Chapter 27. Assault

§ 2701. Simple assault.

§ 2702. Aggravated assault.

§ 2702.1. Assault of law enforcement officer.

§ 2703. Assault by prisoner.

§ 2703.1. Aggravated harassment by prisoner.

§ 2704. Assault by life prisoner.

§ 2705. Recklessly endangering another person.

§ 2706. Terroristic threats.

§ 2707. Propulsion of missiles into an occupied vehicle or onto a roadway.

INCIDENTHIERARCHYDESC	OFFENSES	INCIDENTTRACT
HARRASSMENT/THREAT/ATTEMPT/PHY	2702 Aggravated Assault. / 2709(a) Harassment. / 3503.B11 DEF. TRESPASS ACTUAL COMMUNICATION	2205
THEFT BY DECEPTION	3922 Theft by Deception.	1011
DISCHARGE OF FIREARM INTO OCC.STRUCTURE	2707.1 Discharge of a Firearm into Occupied Structure	2609
HARRASSMENT/THREAT/ATTEMPT/PHY	2709(a)(3) Harassment No Legitimate Purpose	201
PROP MISSILE INTO OCC VEHICLE/OR ROADWAY	2705 Reckless Endangering Another Person. / 3304 Criminal Mischief. / 2707.1 Discharge of a Firearm into Occupied Structure	2609
PUBLIC DRUNKENNESS	5505 Public Drunkenness / 3503.B11 DEF. TRESPASS ACTUAL COMMUNICATION	1702
AGG ASSAULT/OTHER/INJURY	2702 Aggravated Assault.	2020

Project narrative

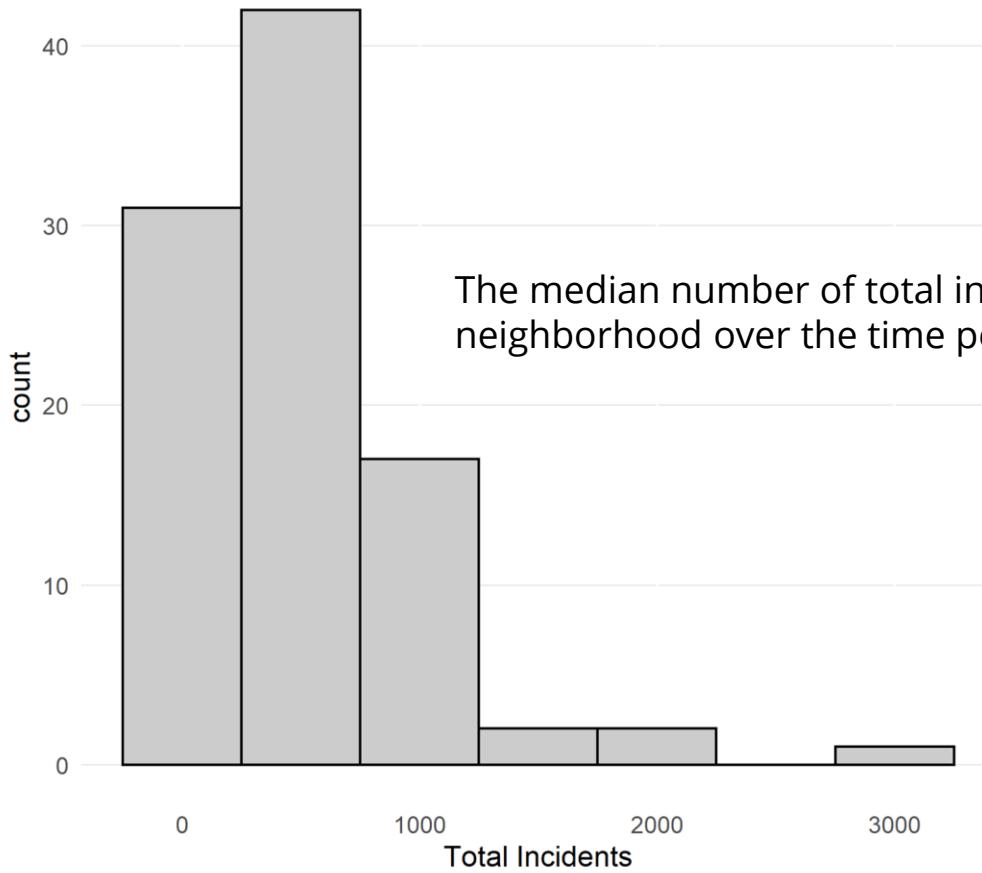
- How do type and frequency of criminal incidents in Pittsburgh change with time and location?
 - How did the pandemic affect crime in Pittsburgh?
 - i.e. Do we observe any significant changes around March 2020?
- What ages and races of people are arrested in Pittsburgh?

Exploratory Data Analysis

Part 1

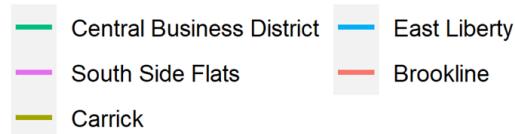
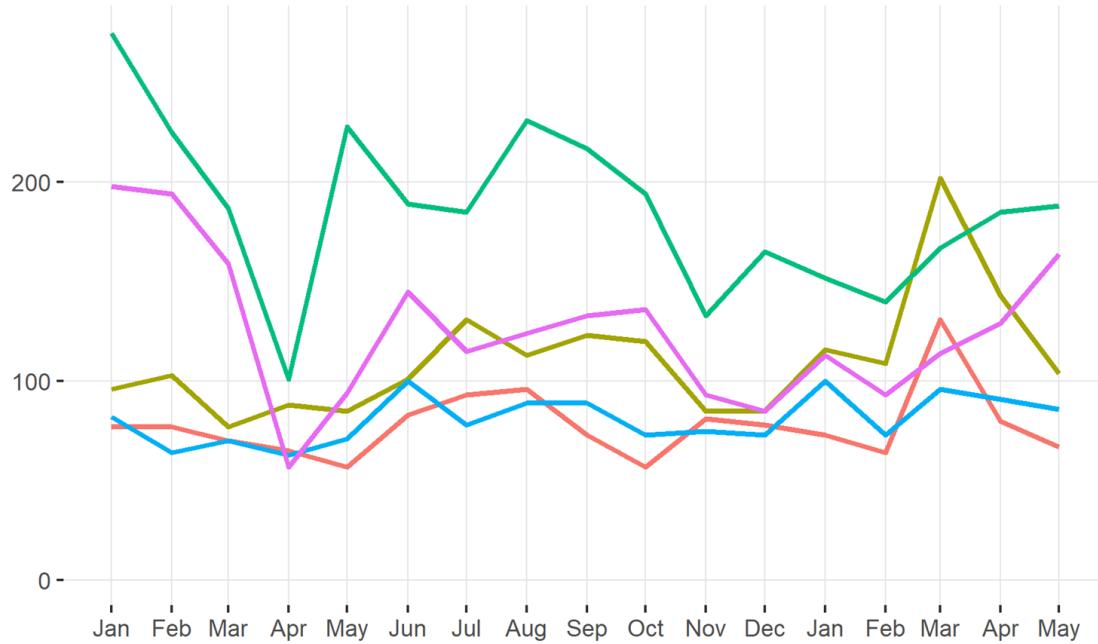
Distribution of Total Incidents per Neighborhood

Jan 2020 - May 2021



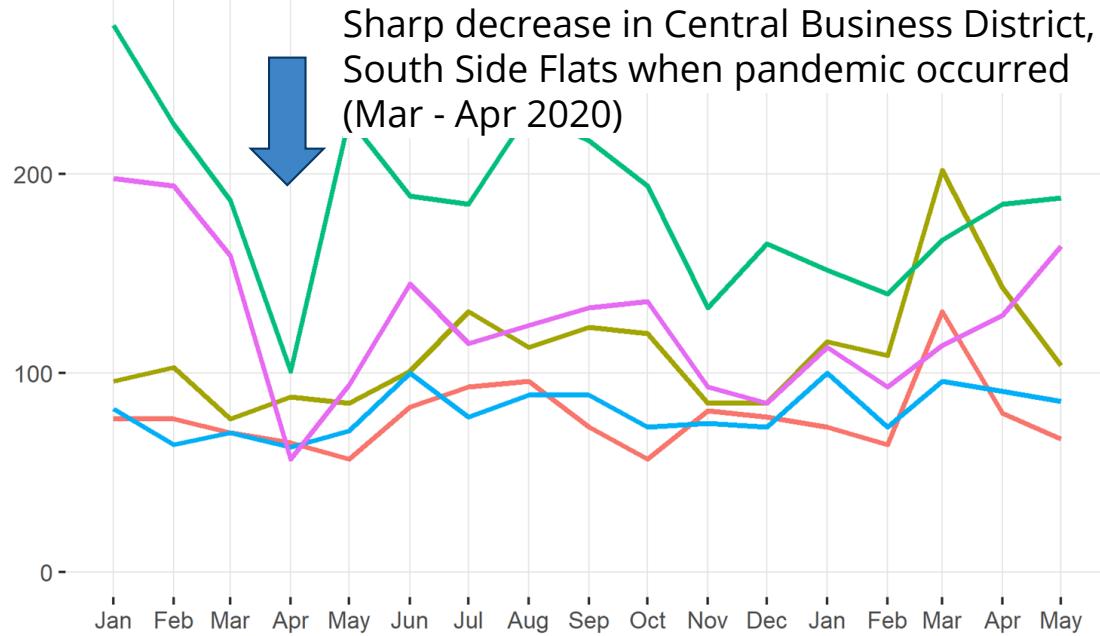
Incidents over Time of Neighborhoods with Top 5 Total Incidents

Jan 2020 - May 2021



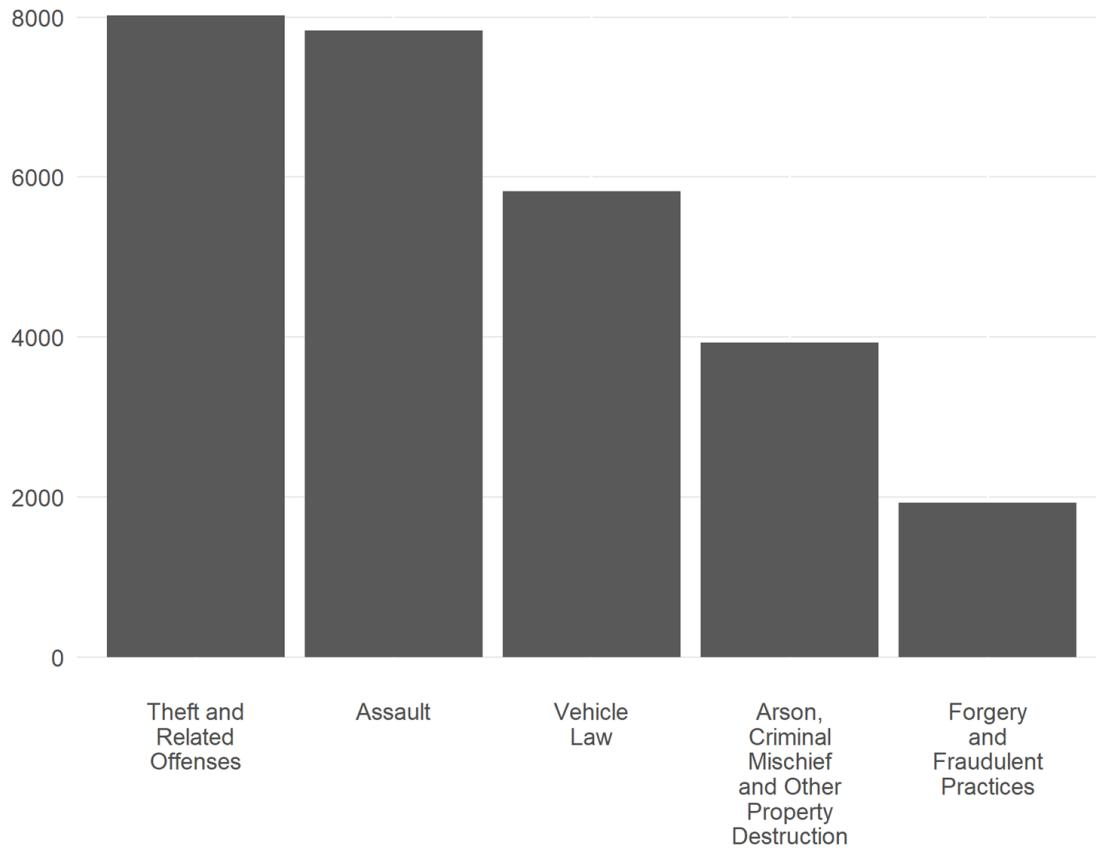
Incidents over Time of Neighborhoods with Top 5 Total Incidents

Jan 2020 - May 2021



Most Frequent Type of Offense by Category

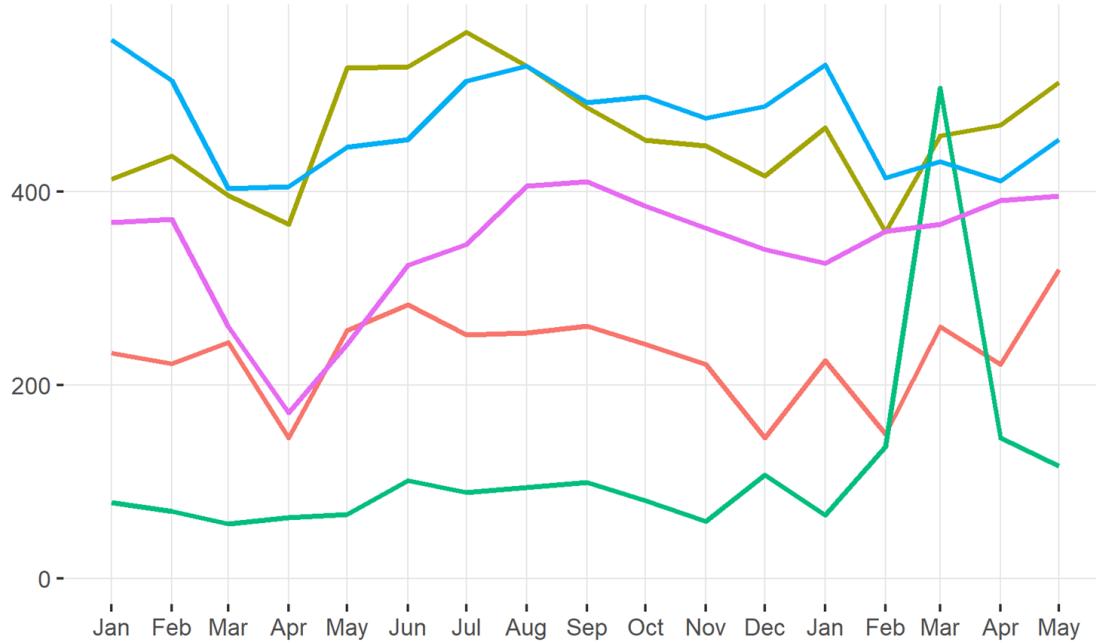
First (Highest Hierarchy) Offense Only



Incidents over Time by Category

Top 5 Categories

Jan 2020 - May 2021



Legend:

- Theft and Related Offenses
- Assault
- Vehicle Law
- Arson, Criminal Mischief and Other Property Destruction
- Forgery and Fraudulent Practices

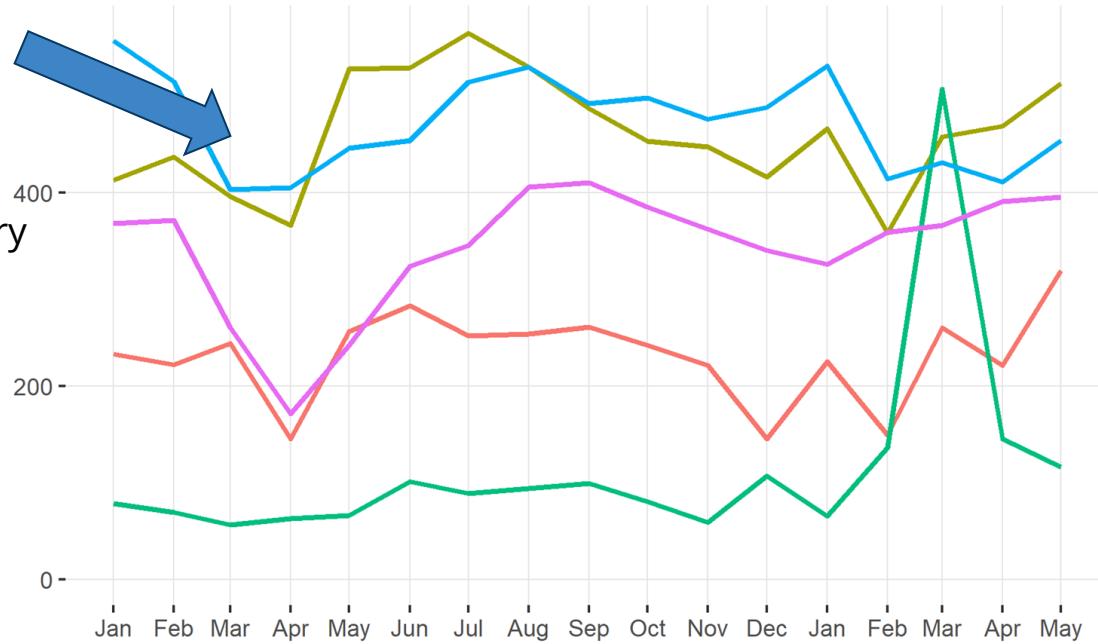
Incidents over Time by Category

Top 5 Categories

Jan 2020 - May 2021

Assault, Vehicle, Arson offenses decreased from March to April 2020

Theft and Forgery did not



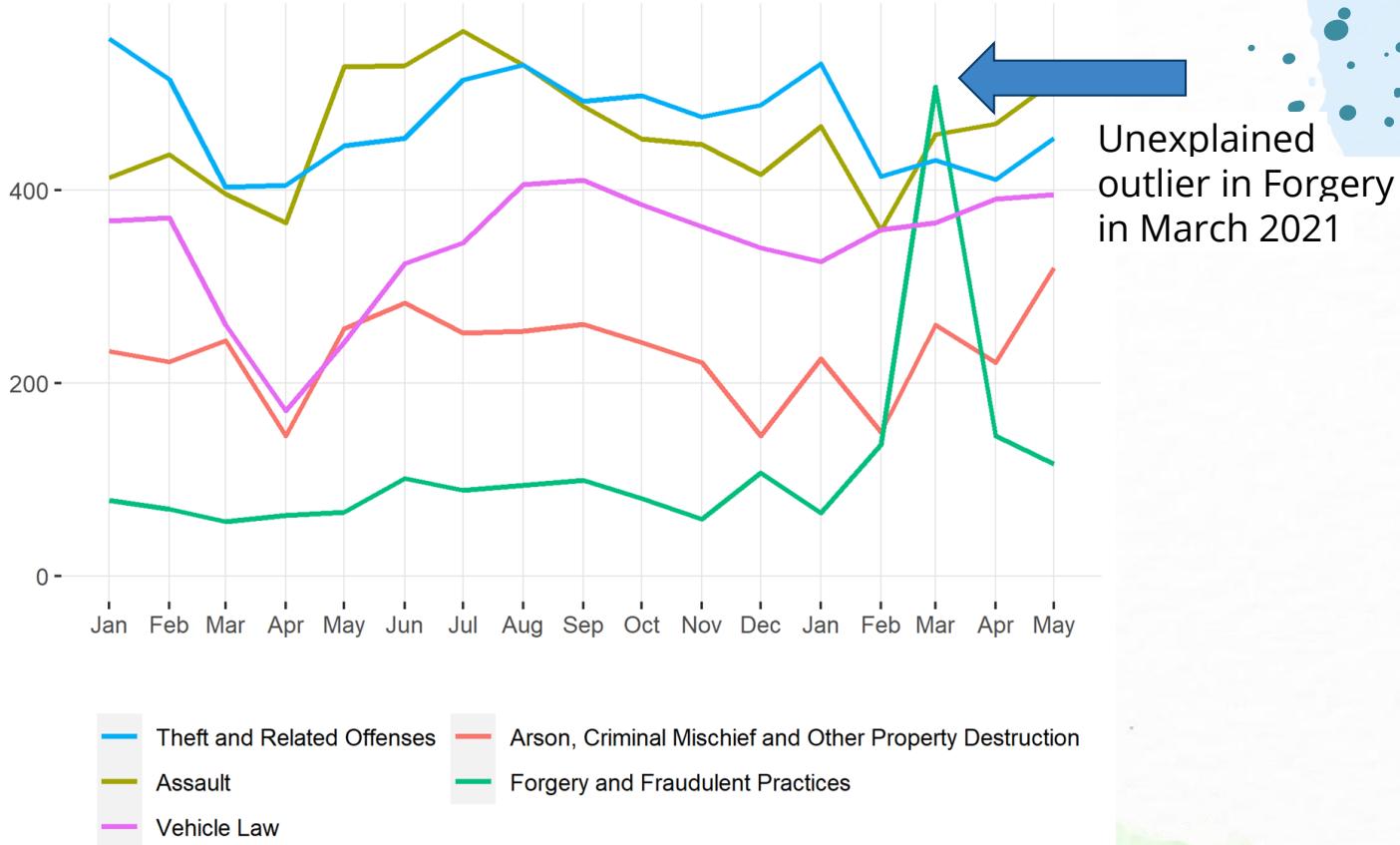
Legend:

- Theft and Related Offenses
- Assault
- Vehicle Law
- Arson, Criminal Mischief and Other Property Destruction
- Forgery and Fraudulent Practices

Incidents over Time by Category

Top 5 Categories

Jan 2020 - May 2021



Victims of unemployment fraud in Pa. blindsided by tax documents

Thousands of apparent victims of Pandemic Unemployment Assistance fraud have received unexpected tax documents reporting income to the IRS they never took.



The Department of Labor & Industry is holding its first budget at the state's House of Representatives on Tuesday, March 2, 2021.
Jelani Splawn / For Spotlight PA

by Rebecca Moss
Published Mar 2, 2021

Hundreds in the Pittsburgh area targeted with fraudulent Chase Bank debit cards, ID theft



HALLIE LAUER
Pittsburgh Post-Gazette

MAR 10, 2021 10:17 AM

3 Pittsburghers accused of defrauding state of nearly \$1M in pandemic unemployment assistance



PITTSBURGH POST-GAZETTE
Got a news tip? 412-263-1601
localnews@post-gazette.com

MAR 29, 2021 2:27 PM

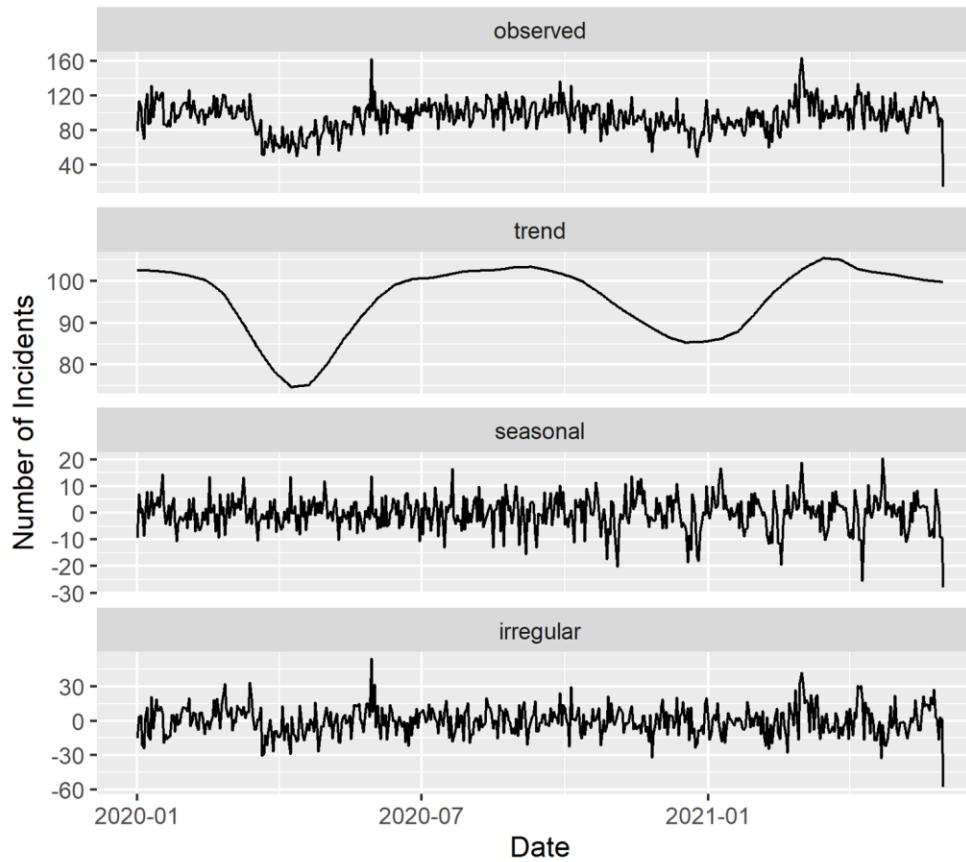
Visualizations

Pa

Seasonal Decomposition of Criminal Offense Incidents

Jan 2020 - May 2021

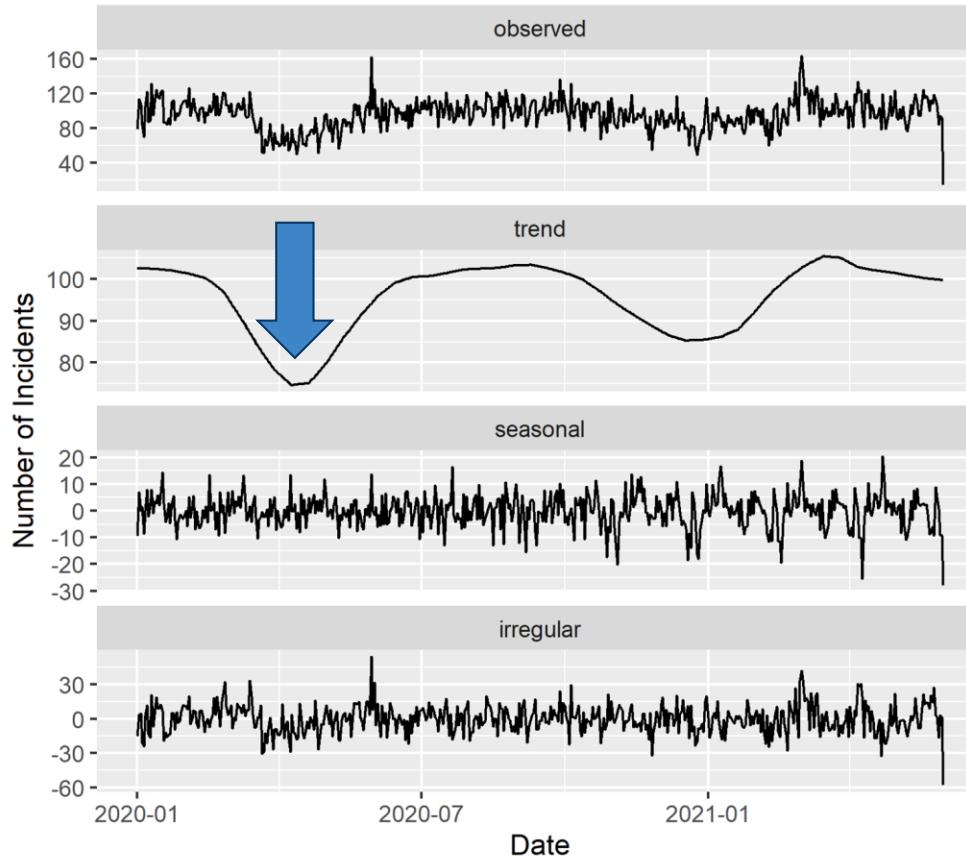
Weekly frequency

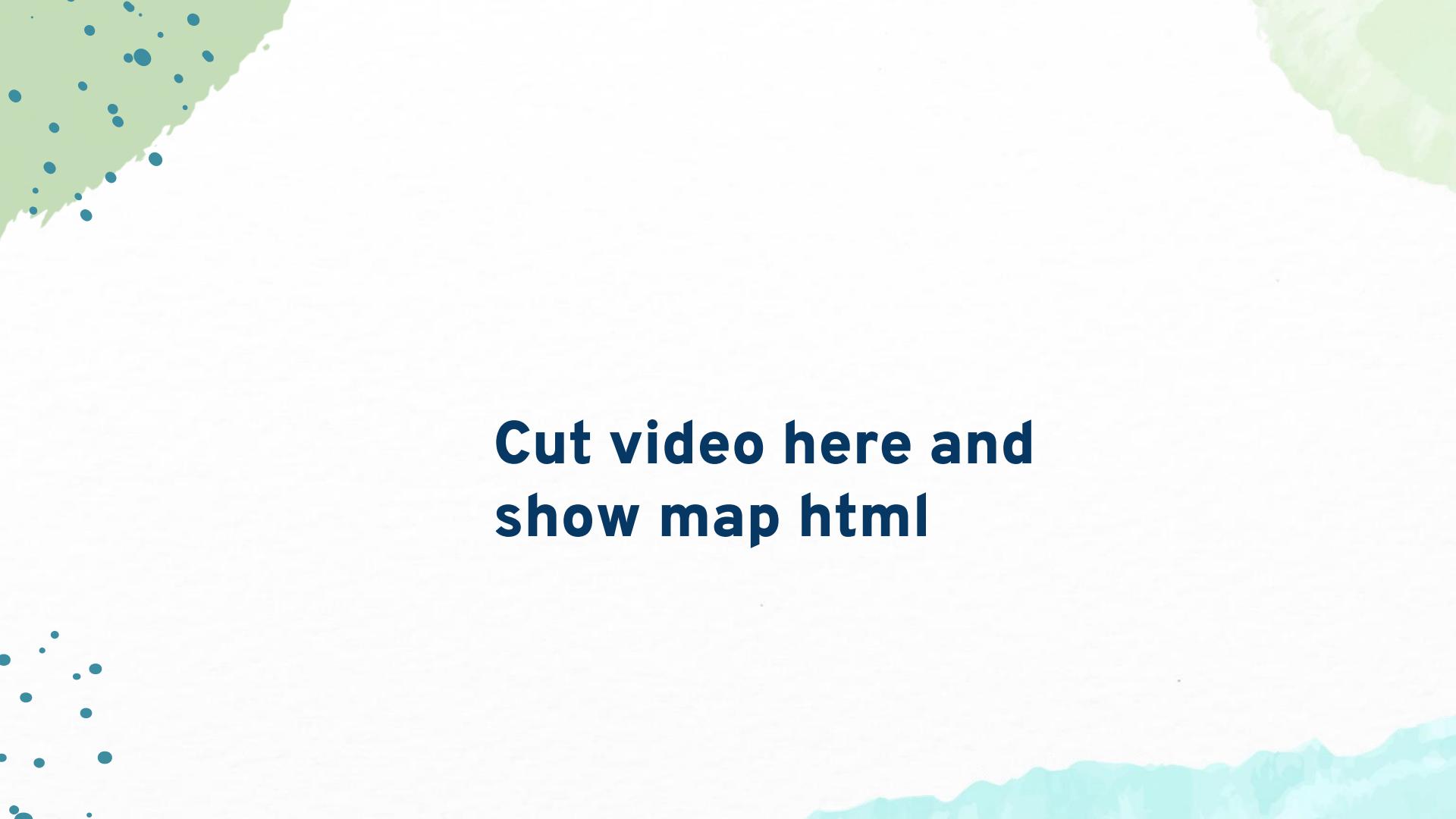


Seasonal Decomposition of Criminal Offense Incidents

Jan 2020 - May 2021

Weekly frequency

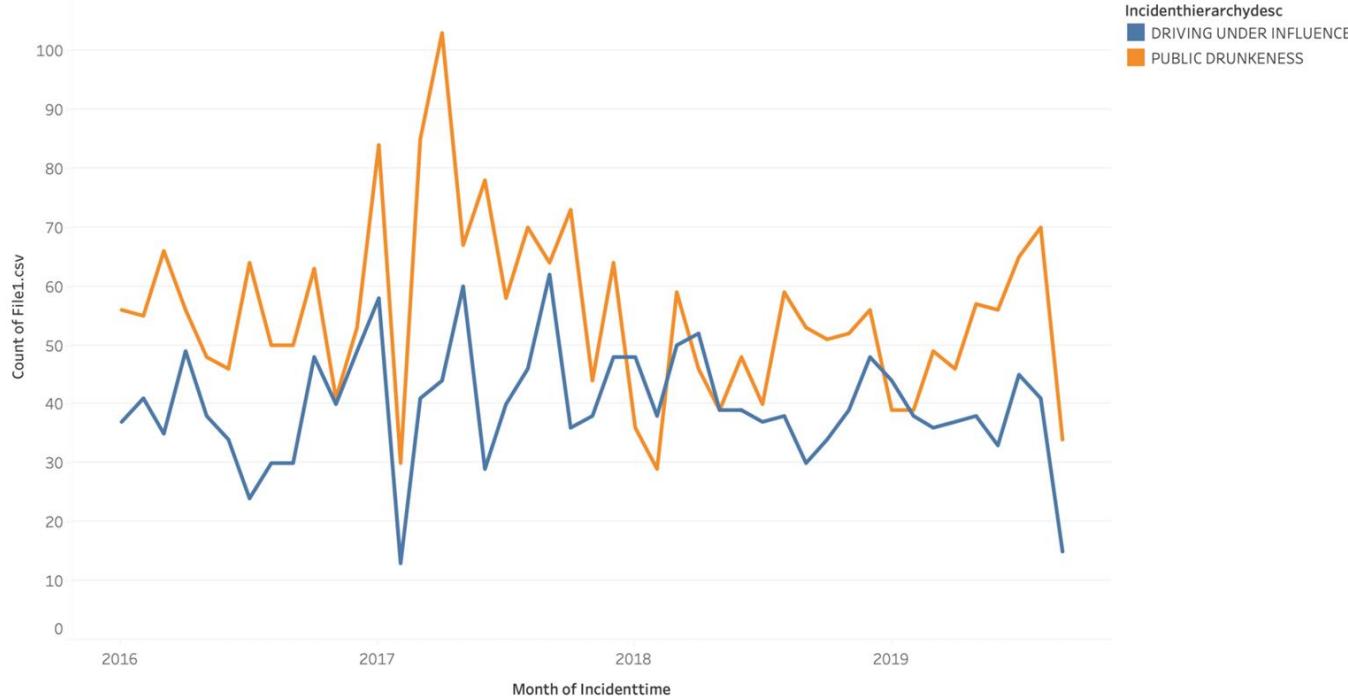




**Cut video here and
show map html**

Visualization #3

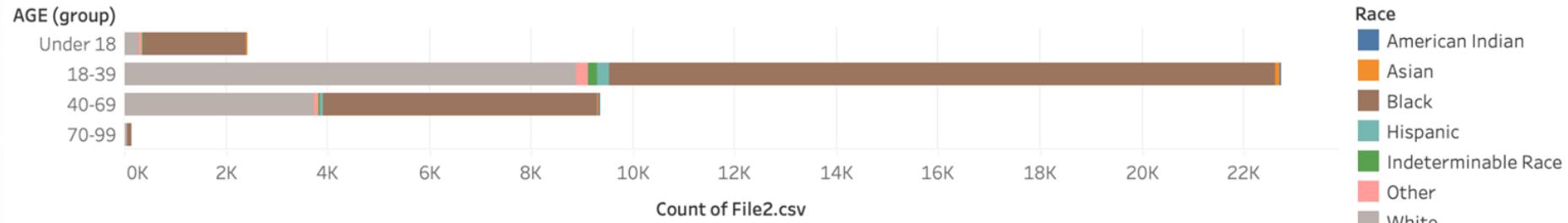
line graph driving under influence vs public drunkenness over time



The trend of count of File1.csv for Incidenttime Month. Color shows details about Incidenthierarchydesc. The data is filtered on Incidenttime, which includes dates on or after 4/14/2014 12:00:00 AM. The view is filtered on Incidenthierarchydesc, which keeps DRIVING UNDER INFLUENCE and PUBLIC DRUNKENNESS.

Visualization #4

stacked bar chart for age and race



Count of File2.csv for each AGE (group). Color shows details about Race. The view is filtered on AGE (group), which keeps 18-39, 40-69, 70-99 and Under 18.

Data Management Tool

- Github
 - Provides a Web-based graphical interface
 - Provides access control and several collaboration features
 - wikis and basic task management tools for every project
- Uses:
 - Store extra information
 - Keep copies of our visualizations
 - Store all DMPs

Repository

- Kilthub
 - Retained for 10 years
 - After 10 years...
 - Kilthub admin determines if there's a need for continued storage of data
- Restrictions
 - No offensive material, false statements, or information that I do not have authority to share
 - No materials supporting criminal activity, materials that invade privacy, materials that are embarrassing to CMU's Platform Provider, or junk mail

Ethical Considerations

- Privacy of criminals
 - Type of crime
- Racial discrimination
- Socio-economic discrimination

Documentation

- Document data source, file type, resources to use, etc.
- Brainstorm in Google Doc
- Write our own individual parts
 - 5 categories in DMPtool
- Put it into DMPtool



Thank you!