



Safe in Seattle: Crime trend analysis

Data 557

Winter 2019

Ben Brodeur Mathieu, Lauren Heintz, Will Wright, Tara Wilson



Study data

Occurred date	Time	Region	Neighborhood	Type of crime
2009-01-01	0800	NE	UNIVERSITY DISTRICT	BURGLARY-RESIDENTIAL
...
2018-12-31	2330	E	CAPITOL HILL	HOMICIDE

Seasonality & time of day



Hypotheses

Null Hypotheses:

There is no difference in volume of crime based on hour of day, time of day, month of year, and season.

Alternative Hypotheses:

There is a difference in volume of crime based on hour of day, time of day, month of year, and season.



Definitions of data

‘Time of Day’

Morning	6am - 12pm
Afternoon	12pm - 6pm
Evening	6pm - 12am
Night	12am - 6am

‘Season’

Winter	December, January, February
Spring	March, April, May
Summer	June, July, August
Fall	September, October, November



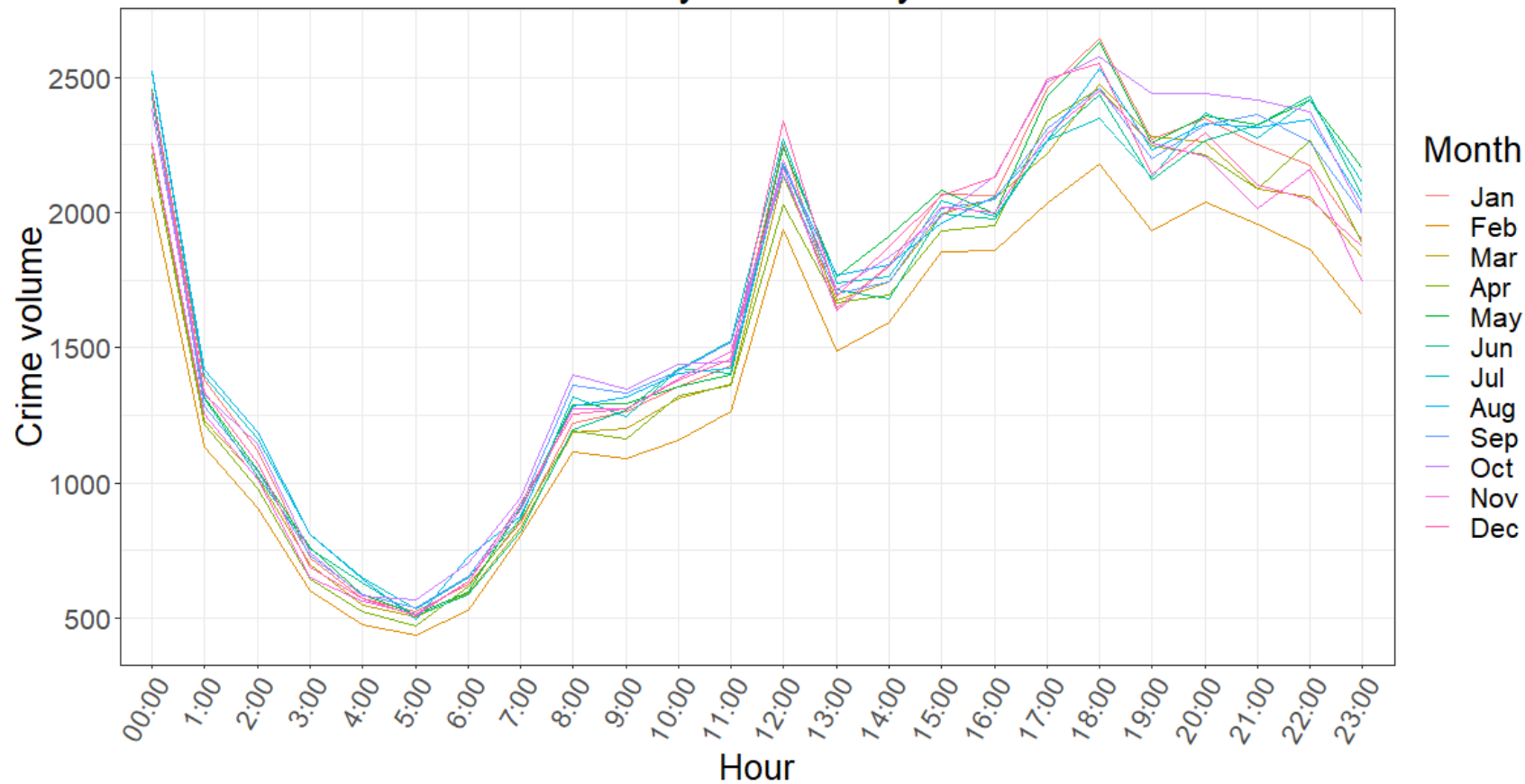
Statistical methods

ANOVA test of independence

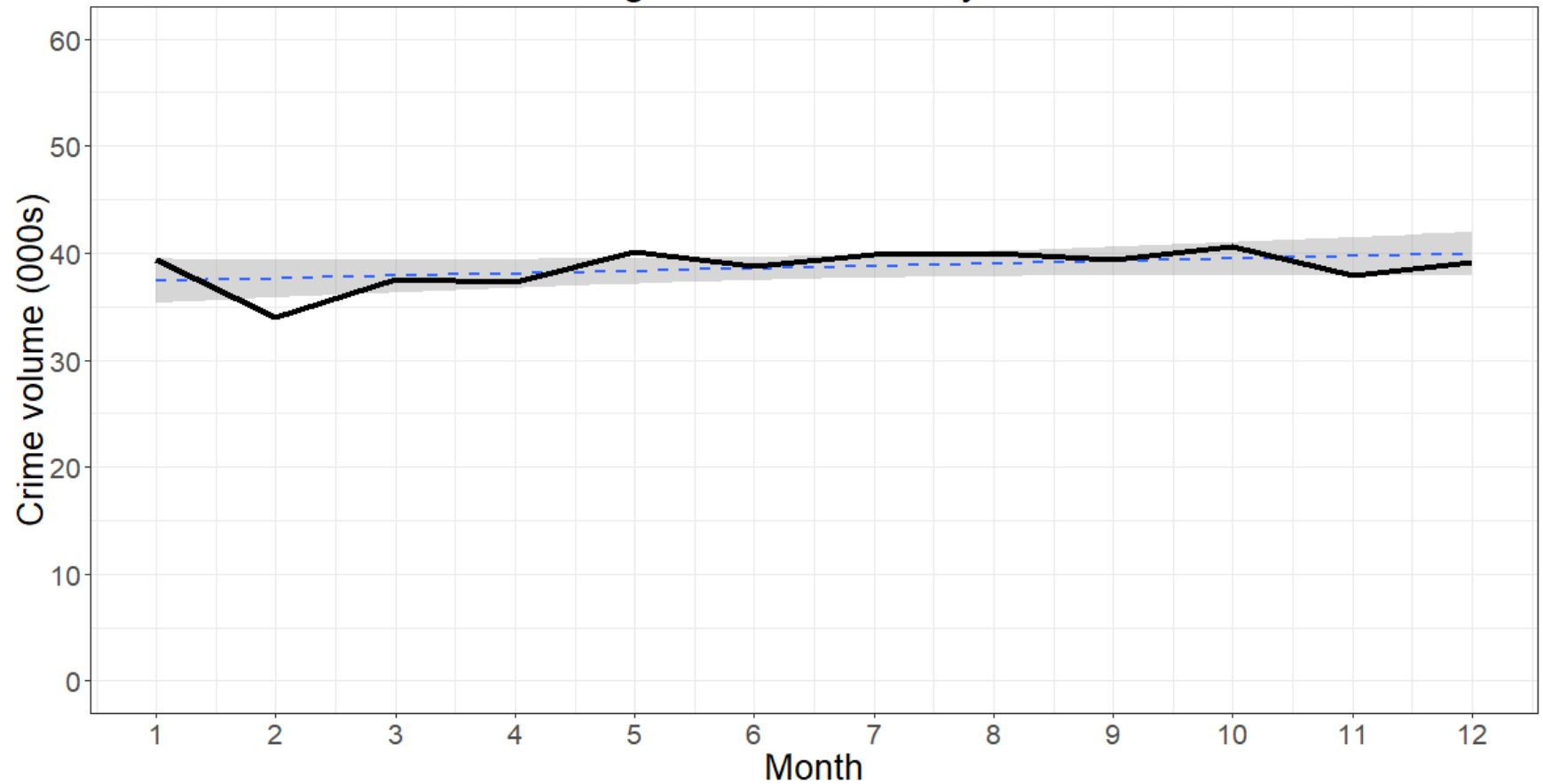
Assumptions:

- Independence
- Constant variance of residuals
- Normality or large sample size

Crime volume by hour of day and month



Average crime volume by month





Results

Metric	Formula	Pr(>F)
Time of Day	<code>aov(crimeVolume ~ factor(TimeOfDay),data)</code>	3.46 e-13
Hour	<code>aov(crimeVolume ~ factor(Hour),data)</code>	< 2.0 e-16
Season	<code>aov(crimeVolume ~ factor(Season),data)</code>	0.997
Month	<code>aov(crimeVolume ~ factor(Month),data)</code>	0.979

Median income



Hypotheses

Null Hypothesis:

There is no difference in volume of crime based on median neighborhood income.

Alternative Hypothesis:

There is a difference in volume of crime based on median neighborhood income.



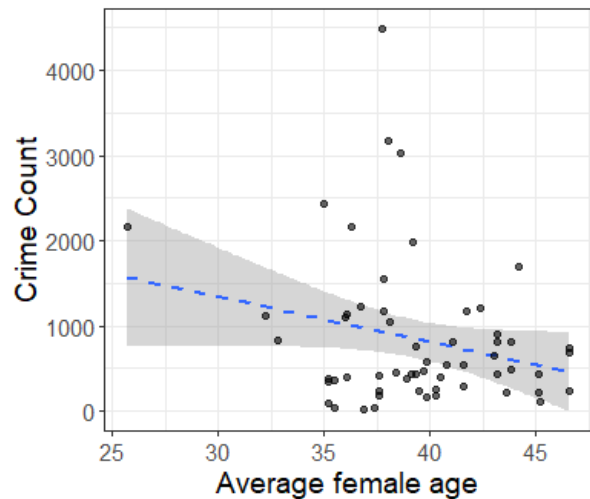
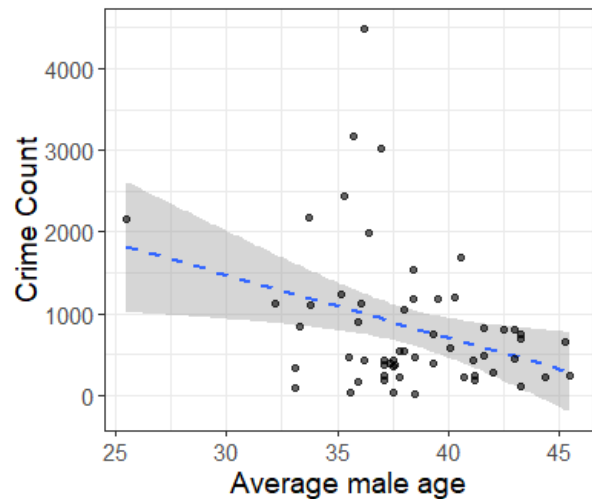
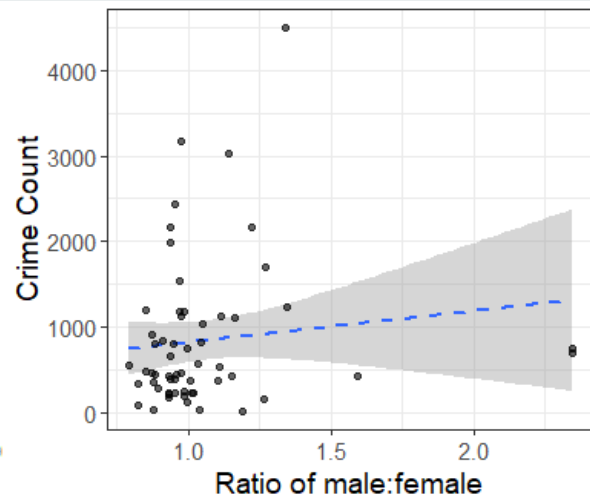
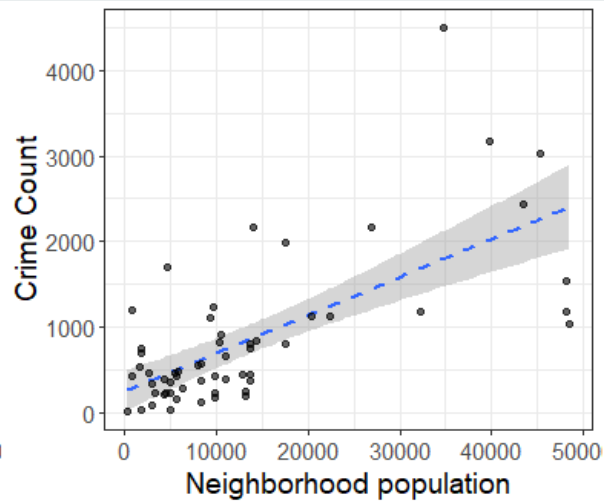
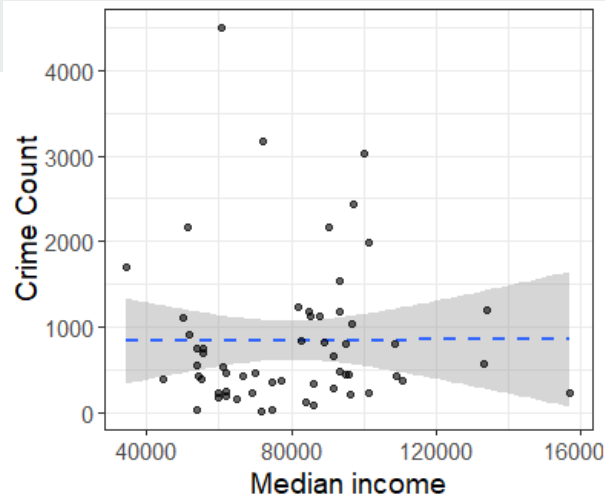
Neighborhood demographic data

Population

Median income

Average age for males and females

Ratio of genders





Statistical methods

Linear regression

Assumptions:

- Independence
- Linearity
- Normality or large sample size
- Constant variance of residuals
 - Not met → use robust standard errors



Linear regression model output

	Estimate	Robust p-value
Median income	9.85e-04	0.810
Population	4.56e-02	< 0.001
Ratio male to female	6.02e+02	0.040
Average age (M)	-1.28e+02	0.001
Average age (F)	8.33e+01	0.011

Seattle regions vs. crime types



Definitions of data

Regions: defined by the team based on the SPD neighborhoods

Crimes:

- Car Prowl
- Bicycle theft
- Motor vehicle theft
- Robbery - street
- Burglary - residential





Hypothesis

Null Hypothesis:

There is no relationship between type of crime and region of Seattle.

Alternative Hypothesis:

There is a relationship between type of crime and region of Seattle.



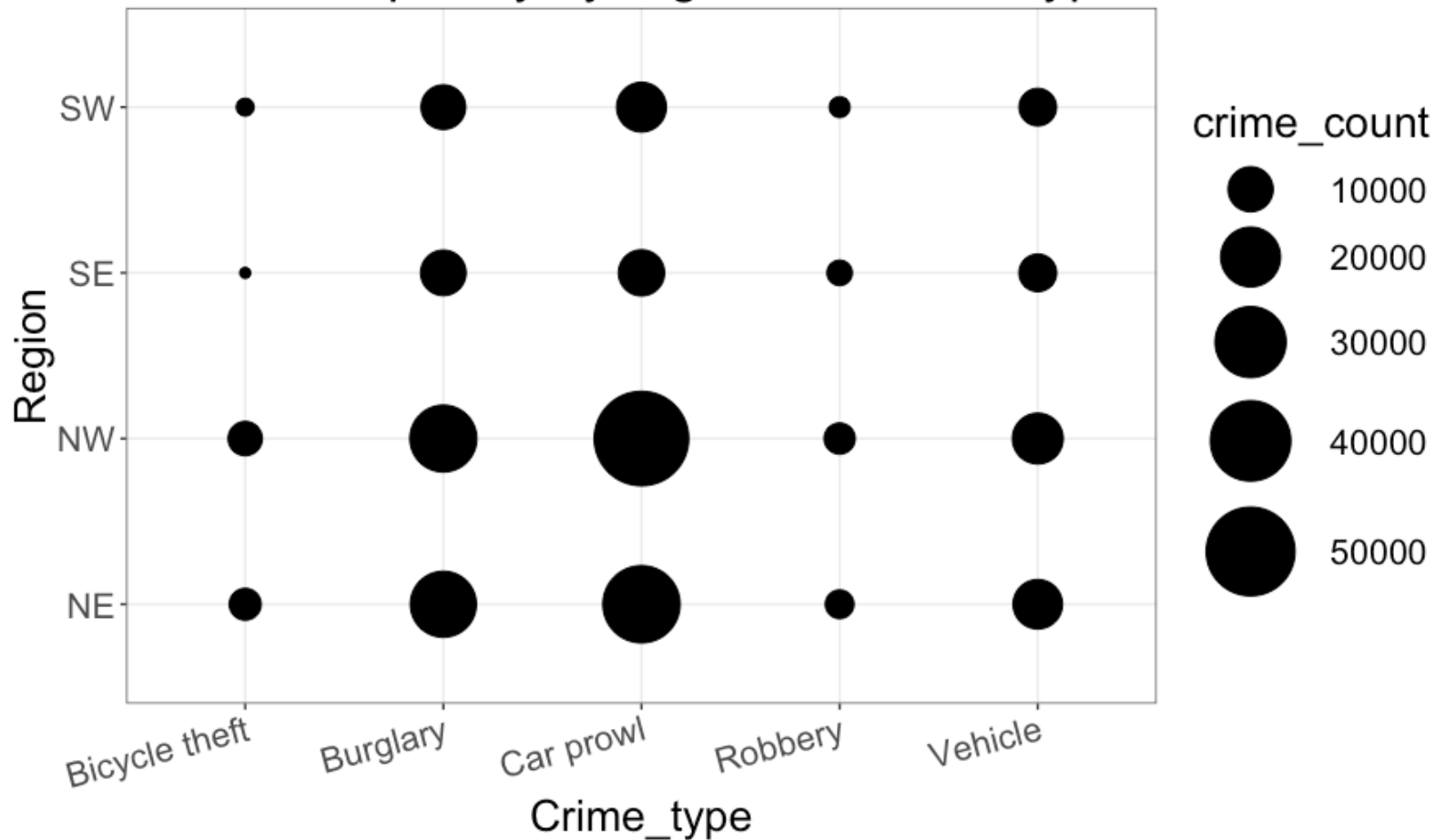
Statistical methods

Chi-squared test of independence

Assumptions:

- Independent samples
- Approximately chi-square (Overall sample size, Cell count)

Crime frequency by region and crime type



Questions
