**Project Proposal**

Submit the completed proposal to the D2L “Project Proposal” Dropbox

**Group Members:­­­­­­­­­­­­­­­­**

Tuli Kariyapperuma

Vasu Chouhan

Amir Targholizadeh

Jaspreet Singh

Curtis McAndrew

What is the general topic of this study?

How have crime rates in Vancouver for the past sixteen years been affected by time of day, type of crime, and where the crime was committed?

1. Describe an individual and the population for your study. Be very specific.

Population: The ~600,000 crimes registered by the Vancouver Police Department, as collected from the City of Vancouver Open Data Catalogue and uploaded to Kaggle: <https://www.kaggle.com/agilesifaka/vancouver-crime-report/version/1I>; limiting our selection to dates between January 1st 2004 and December 31st 2018 (to have full calendar years in our population).

Individual: A single reported crime from this list.

1. What sample size will you use? n = 60,000 if Simple Random Sampling is used; 10% if Stratified is used.
2. What variables will you consider? i.e. what questions will you ask each individual? (Consider using 5-8 variables). Classify each variable as quantitative or categorical. For categorical variables, state the categories (eg. If the variable is Faculty that a student is in, then the categories might be: Arts, Science, Career, Other.) I suggest that you consider no more than five categories for each categorical variable. *Note: It is a good idea to include a variety of categorical and quantitative variables.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Variable Description** | **Variable Class** | **Possible Values of Variable** |
| Year | Month and year crime was reported | Interval | Calendar year, from 2004 to 2018 |
| Time of Day | Hour and minute crime was reported | Interval | One of eight 3-hour blocks, starting from 00:01 to 03:00, etc. |
| Type of Crime | Category allocated to the reported crime | Nominal | Break and Enter  Homicide  Mischief  Offence Against a Person  Theft  Vehicle Collision or Pedestrian Struck |
| Neighbourhood | Location of the reported crime  (Note: we will likely choose a method to combine neighbourhoods, to reduce the quantity of categories. If so, the exact method used will be determined as we proceed with the project.) | Nominal | Central Business District  South Cambie  Stanley Park  Oakridge  Kensington-Cedar Cottage  Fairview  Strathcona  Grandview-Woodland  Kitsilano  Hastings-Sunrise  West End  Marpole  Mount Pleasant  Sunset  Killarney  Riley Park  Renfrew-Collingwood  Arbutus Ridge  Shaughnessy  Kerrisdale  West Point Grey  Dunbar-Southlands  Victoria-Fraserview  Musqueam |

1. How will the individuals be contacted? (personal interview, phone, direct observation, questionnaire, etc.) If you are using a questionnaire, attach a draft of your questionnaire to this proposal.

Not applicable: we are using a pre-existing data set, not contacting any individuals.

1. State all objectives that require univariate and bivariate data analyses (you must have at least 3 univariate objectives and 2 bivariate objectives).

Univariate:

--How have the quantity of reported crimes changed over the years?

--Which periods of each day tend to have the most total reported crimes?

--Which types of crimes have been the most commonly reported?

--How many crimes have been reported in total over the population’s period for each neighbourhood?

Bivariate:

--How does the time of day affect the type of crime reported?

--How does the neighbourhood affect the time (of day) the crime was reported?

--How does the neighbourhood affect the type of crime reported?

1. State the sampling design that you will use. Describe, step-by-step, how you will select the individuals. Are there any limitations or biases?

We have not firmly decided our method, but it will be one of the following options. We will examine the data in greater detail, and choose the most appropriate method based on our findings.

1. Simple Random Sampling: use R to pull 60,000 random samples from the population.
2. Stratified – Neighbourhood: use R to pull a proportionate random sample of 10% of the reported crimes from each neighbourhood.
3. Stratified – Type of Crime: use R to pull a proportionate random sample of 10% of the reported crimes from each type of crime reported.