

# How Might Vancouver's Trees Face Climate Change?

An analysis of neighbourhoods' trees climate change risk level and resilience

Marina Galvao - August 11th, 2025

## Introduction

It is well known that as climate change progresses, urban areas will start to face ever growing environmental challenges. Notable occurrences which are predicted to increase include wildfire events, as well as negative environmental impacts from species imbalance, such as invasive species causing significant harm to native species.

Nowadays, we have been experiencing increasing wildfire rates on a yearly basis. The implications this poses to different urban areas are not yet fully understood. Increases in wildfires are a consequence of a warming climate, and climate change creates conditions where invasive species can further degrade ecosystems. This is especially important in the context of trees in the genus *Pinus*, as these are known to be highly flammable and a large fuel source for wildfire spread. These risks are only predicted to increase over time as climate change progresses. On another note, the emerald ash borer, an invasive beetle species whose larvae use trees of the genus *Fraxinus* as host, has made its way to Vancouver, and created significant negative repercussions in the flora around Vancouver.

The questions this analysis will investigate are: In the context of climate change,

1. What neighbourhood around Vancouver is most susceptible to be negatively impacted by the emerald ash borer and uncontrollable (highly flammable) wildfires?
2. What understanding do we have of these at-risk neighbourhoods (i.e. biodiversity, tree diameter...) that may help us infer how they might respond to these impacts?
3. How do these two at-risk neighbourhoods compare to all neighbourhoods in Vancouver in terms of biodiversity? Is their biodiversity higher in a way that might imply increased ecosystem resilience against invasive species and wildfire?
4. What might tree height across all different neighbourhoods show about potential crown fire spread (one of the most severe types of wildfire - where wildfire travels from one tree top to another) in Vancouver?

By better understanding neighbourhoods susceptibility to these already significant and increasing risks, work can be done around ensuring these neighbourhoods get the attention they deserve, and that appropriate action can be taken prior to negative impacts increasing. Examples may be installing emerald ash borer traps in the neighbourhoods highlighted as most susceptible prior to areas that are less at risk, or removing flammable species and increasing diversity in neighbourhoods with a large number of fire-prone trees. Moreover, understanding potential resilience levels is just as important as understanding and interpreting potential risks in order to comprehend these complex ecosystems; for this reason, resilience will also be investigated in this analysis.

To answer the questions presented here, data from a public trees dataset containing information about a variety of factors related to trees around the city of Vancouver will be used. The data was obtained from the city of Vancouver's [Open Data Portal](#), and [cleaned up by the course instructors](#) for use. The tagged trees CSV data contains multiple rows where each row represents one tree. The data contains quantitative columns such as tree diameter, but also many other features including the exact species, its date planted, the neighbourhood and street which it is located on, amongst other important characteristics.

## Analysis

In [1]: # Required packages

```
import altair as alt
import pandas as pd
```

Table 1 - Vancouver Trees Dataframe

In [2]: # Importing the data

```
url = "https://raw.githubusercontent.com/UBC-MDS/data_viz_wrangled/main/data/Trees_data_sets/small_unique_vancouver.csv"
van_trees = pd.read_csv(url)
van_trees
```

Out[2]:

	Unnamed: 0	std_street	on_street	species_name	neighbourhood_name	date_planted	diameter	street_side_name	genus_name	assigned	...	plant_area	...
0	10747	W 20TH AV	W 20TH AV	PLATANOIDES	Riley Park	2000-02-23	28.5	EVEN	ACER	N	...	15	
1	12573	W 18TH AV	W 18TH AV	CALLERYANA	Arbutus-Ridge	1992-02-04	6.0	ODD	PYRUS	N	...	7	
2	29676	ROSS ST	ROSS ST	NIGRA	Sunset		NaN	12.0	ODD	PINUS	N	...	7
3	8856	DOMAN ST	DOMAN ST	AMERICANA	Killarney	1999-11-12	11.0	EVEN	FRAXINUS	N	...	7	
4	21098	EAST BOULEVARD	EAST BOULEVARD	HIPPOCASTANUM	Shaughnessy		NaN	15.5	ODD	AESCULUS	Y	...	N
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4995	6132	E 53RD AV	E 53RD AV	SERRULATA	Victoria-Fraserview		NaN	17.0	EVEN	PRUNUS	N	...	9
4996	5642	E 32ND AV	E 32ND AV	XX	Kensington-Cedar Cottage	2014-01-14	3.0	EVEN	CORNUS	N	...	10	
4997	8777	DAWSON ST	DAWSON ST	TULIPIFERA	Killarney	2002-04-15	3.5	EVEN	LIRIODENDRON	N	...	7	
4998	23489	E 13TH AV	E 13TH AV	INVOLUCRATA	Mount Pleasant	2003-12-02	5.5	EVEN	DAVIDIA	N	...	5	
4999	7450	CULLODEN ST	CULLODEN ST	CAMPESTRE	Kensington-Cedar Cottage		NaN	3.0	ODD	ACER	N	...	8

5000 rows x 21 columns

Observing Table 1, it is evident that each row represents a different tree recorded and its respective characteristics. There are 5000 rows total and 21 columns in the dataframe.

Table 2 – Description of Dataframe

In [3]: `van_trees.describe()`

Out[3]:	Unnamed: 0	diameter	civic_number	tree_id	height_range_id	on_street_block	latitude	longitude
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	14861.920400	12.340888	2975.707600	128682.584600	2.73440	2960.227000	49.247349	-123.107128
std	8680.023278	9.266600	2078.580429	75412.260406	1.56957	2086.861052	0.021251	0.049137
min	2.000000	0.000000	2.000000	36.000000	0.00000	0.000000	49.202783	-123.220560
25%	7192.750000	4.000000	1300.500000	61321.500000	2.00000	1300.000000	49.230152	-123.144178
50%	14870.000000	10.000000	2639.000000	130130.500000	2.00000	2600.000000	49.247981	-123.105861
75%	22366.750000	18.000000	4123.000000	191332.000000	4.00000	4100.000000	49.263275	-123.063484
max	29992.000000	71.000000	9113.000000	270750.000000	9.00000	9100.000000	49.293930	-123.023311

The summary table shows that the count for the different variables present is 5000, the same as the number of rows present (Table 1). The highest diameter observed is 71cm, and the highest category of height\_range\_id is 9. The mean diameter observed is 12.34cm, and mean height\_range\_id category is 2.73. The highest standard deviations can be observed in the unnamed and tree\_id variables, as these values do not have any distinct relationship in the data collected, and merely represent ID numbers.

In [4]: `van_trees.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   Unnamed: 0        5000 non-null   int64  
 1   std_street       5000 non-null   object  
 2   on_street        5000 non-null   object  
 3   species_name     5000 non-null   object  
 4   neighbourhood_name 5000 non-null   object  
 5   date_planted    2363 non-null   object  
 6   diameter         5000 non-null   float64 
 7   street_side_name 5000 non-null   object  
 8   genus_name       5000 non-null   object  
 9   assigned          5000 non-null   object  
 10  civic_number     5000 non-null   int64  
 11  plant_area       4950 non-null   object  
 12  curb              5000 non-null   object  
 13  tree_id          5000 non-null   int64  
 14  common_name      5000 non-null   object  
 15  height_range_id  5000 non-null   int64  
 16  on_street_block  5000 non-null   int64  
 17  cultivar_name    2658 non-null   object  
 18  root_barrier     5000 non-null   object  
 19  latitude          5000 non-null   float64 
 20  longitude         5000 non-null   float64 
dtypes: float64(3), int64(5), object(13)
memory usage: 820.4+ KB
```

The columns used in this analysis will be: "neighbourhood\_name", "genus\_name", "height\_range\_ID", and "diameter". The column "neighbourhood\_name" contains the name of the exact neighbourhood in Vancouver where the tree is located. The "genus\_name" represents the genus that the tree belongs to, with the genus being a category in tree taxonomy. We will also be using the "height\_range\_ID", which is a categorical value based on tree height, created using the following: 0 = 0-10 ft, 1 = 10-20 ft, 2 = 20-30 ft, and 10 = 100+ ft. Lastly, "diameter" stands for Diameter at Breast Height (DBH), a measure of the diameter of the tree at approximately 4.5 feet off the ground, often used in the forestry industry as an important data measure.

The columns will help us answer questions as described below:

1. What neighbourhood around Vancouver is most susceptible to be negatively impacted by the emerald ash borer and uncontrollable (highly flammable) wildfires? --> Can be analyzed using "neighbourhood\_name" and "genus"
2. What understanding do we have of these at-risk neighbourhoods (i.e. biodiversity, tree diameter...) that may help us infer how they might respond to these impacts? --> Can be analyzed using "neighbourhood\_name", "genus\_name", and "diameter"
3. How do these two at-risk neighbourhoods compare to all neighbourhoods in Vancouver in terms of biodiversity? Is their biodiversity higher in a way that might imply increased ecosystem resilience against invasive species and wildfire? --> Can be analyzed using "neighbourhood\_name" and "genus\_name"
4. What might tree height across all different neighbourhoods show about potential crown fire spread (one of the most severe types of wildfire - where wildfire travels from one tree top to another) in Vancouver? --> Can be analyzed using "neighbourhood\_name" and "height\_range\_ID"

Looking at the above summary of the dataframe, it is evident that the columns "date\_planted", "plant\_area", and "cultivar\_name" have null values. Since these are not used in this analysis, null values for these columns will not need to be removed in this case. The other remaining columns will not be required in our analysis, as these do not pertain to our question. As a result, we can keep the dataframe as is.

In [5]: `# Let's investigate the genera values present`

`van_trees['genus_name'].unique()`

```
Out[5]: array(['ACER', 'PYRUS', 'PINUS', 'FRAXINUS', 'AESCULUS', 'PARROTIA',
 'MAGNOLIA', 'QUERCUS', 'MALUS', 'FAGUS', 'PRUNUS', 'AMELANCHIER',
 'GLEIDITSSIA', 'LIRIODENDRON', 'TILIA', 'BETULA', 'CERCIDIOPHYLLUM',
 'CATALPA', 'SORBUS', 'CORNUS', 'METASEQUOIA', 'CARPINUS',
 'CRATAEGUS', 'SYRINGA', 'KOELREUTERIA', 'CHAMAECYPARIS',
 'PLATANUS', 'STYRAX', 'ILEX', 'LIQUIDAMBAR', 'STEWARTIA',
 'HIBISCUS', 'ROBINIA', 'ULMUS', 'PSEUDOTSUGA', 'CELTIS', 'THUJA',
 'JUGLANS', 'CORYLUS', 'ALNUS', 'SEQUOIADENDRON', 'LABURNUM',
 'GINKGO', 'ABIES', 'CEDRUS', 'ZELKOVA', 'GYMNOCLADUS', 'NYSSA',
 'PTEROCARPA', 'CERCIS', 'DAVIDIA', 'CASTANEA', 'PICEA', 'SEQUOIA',
 'TSUGA', 'CALOCEDRUS', 'POPULUS', 'CHITALPA', 'MESPILUS',
 'OSTRYA', 'NOTHOFAGUS', 'JUNIPERUS', 'ARAUCARIA', 'SOPHORA',
 'EUCOMMIA', 'PTELEA', 'CLADRASTIS'], dtype=object)
```

```
In [6]: # How many Fraxinus trees are there?
fraxinus_count = (van_trees['genus_name'] == 'FRAXINUS').sum()
fraxinus_count
```

```
Out[6]: np.int64(238)
```

```
In [7]: # How many Pinus trees are there?
pinus_count = (van_trees['genus_name'] == 'PINUS').sum()
pinus_count
```

```
Out[7]: np.int64(22)
```

Filtering the genus\_name column to show only the genera Fraxinus and Pinus will be necessary to answer some of the questions presented here.

Table 3, Filtered dataframe with only the genus Fraxinus and Pinus.

```
In [8]: # New dataframe with only the two genera that we want to investigate
filtered_trees = van_trees[
    (van_trees['genus_name'] == 'FRAXINUS') | (van_trees['genus_name'] == 'PINUS')
]

filtered_trees
```

```
Out[8]:   Unnamed: 0 std_street on_street species_name neighbourhood_name date_planted diameter street_side_name genus_name assigned ... plant_area curb t
  2    29676 ROSS ST ROSS ST      NIGRA           Sunset     NaN     12.0        ODD     PINUS   N ...    7   Y  -
  3    8856  DOMAN ST  DOMAN ST  AMERICANA       Killarney  1999-11-12    11.0        EVEN    FRAXINUS   N ...    7   Y  1
  81   13444 W 13TH AV  YEW ST  AMERICANA      Kitsilano     NaN      8.0        ODD    FRAXINUS   N ...    7   Y  2
  82   29656 E PENDER ST E PENDER ST      ORNUS  Hastings-Sunrise  2004-11-04     4.0        EVEN    FRAXINUS   N ...    9   Y  2
  99   18388 PRIOR ST PRIOR ST    CONTORTA      Strathcona     NaN     13.0        EVEN     PINUS   Y ...   B   Y  2
  ...   ...
  4900  15288 BEATRICE ST BEATRICE ST  PENNSYLVANICA Victoria-Fraserview  2002-02-14    12.0        ODD    FRAXINUS   N ...   10   Y  -
  4930  1714 HARO ST  HARO ST      ORNUS          West End     NaN      7.0        EVEN    FRAXINUS   N ...   10   Y  2
  4957  3498 W 62ND AV W 62ND AV  AMERICANA      Marpole     NaN     16.0        ODD    FRAXINUS   N ...   10   Y  -
  4960  11750 W 15TH AV W 15TH AV  AMERICANA      Fairview     NaN     34.0        EVEN    FRAXINUS   N ...    6   Y  -
  4973  4398 E 60TH AV  E 60TH AV  PENNSYLVANICA      Sunset  2000-02-25     9.0        ODD    FRAXINUS   N ...   10   Y  -
```

260 rows × 21 columns

```
In [9]: # Fraxinus and Pinus neighbourhood graph
genus_chart = alt.Chart(filtered_trees).mark_bar().encode(
    x=alt.X('count():Q', title='Number of Trees'),
    y=alt.Y('neighbourhood_name:N', sort='-x', title='Neighbourhood'),
    color=alt.Color('genus_name:N', title='Genus', scale=alt.Scale(scheme='set2')),
    tooltip=['genus_name:N', 'neighbourhood_name:N', 'count():Q']
).properties(
    title='Fraxinus and Pinus Trees by Neighbourhood'
)

genus_chart
```

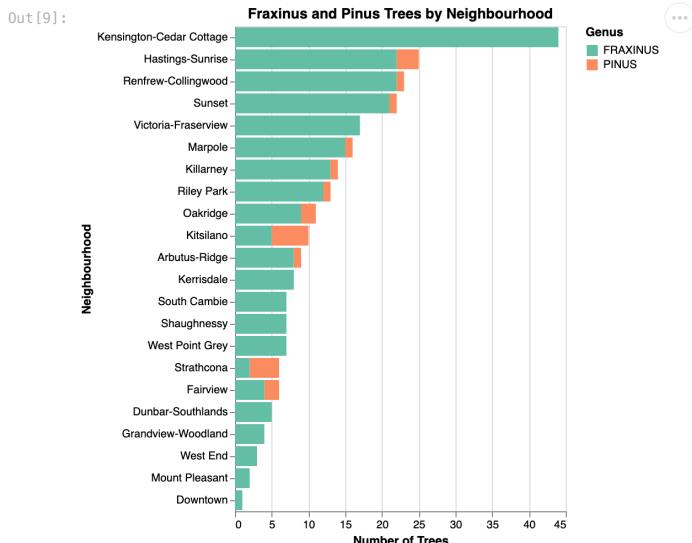


Figure 1 - Number of trees and neighbourhoods graphed in a bar chart. Genera include Fraxinus and Pinus.

Figure 1 shows that Kensington-Cedar Cottage is the neighbourhood with the most trees of the genus Fraxinus, possessing 44 trees of this genus. It is also evident that Kitsilano has the most trees of the genus Pinus – although the highest, this value is of only 5 trees in this genus.

What about the biodiversity in the genera planted in these neighbourhoods? This could help increase ecosystem resilience. Investigating how genera are selected in these neighbourhoods, and how much variety is present will assist in understanding these ecosystems. In order to do so, the original dataframe needs to be filtered to show only the neighbourhoods being investigated at the moment (Kensington-Cedar Cottage and Kitsilano). It is also necessary to group by neighbourhood and genus\_name, and count the number of trees in order to understand the genus breakdown between the number of trees in each genus.

```
In [10]: # Filter to show only Kensington-Cedar Cottage and Kitsilano
top_neigh_df = van_trees[van_trees['neighbourhood_name'].isin(['Kensington-Cedar Cottage', 'Kitsilano'])]

# Group by neighbourhood and genus_name, count the number of rows (trees)
grouped = (
    top_neigh_df.groupby(['neighbourhood_name', 'genus_name'])
    .size()
    .reset_index(name='tree_count')
)

# Sort values
grouped = grouped.sort_values(by='tree_count', ascending=False)

In [11]: # Create scatterplot
scatter = alt.Chart(grouped).mark_circle(size=100).encode(
    x=alt.X('genus_name:N', title='Genus Name', sort=-y),
    y=alt.Y('tree_count:Q', title='Tree Count'),
    color=alt.Color('neighbourhood_name:N', title='Neighbourhood'),
    tooltip=['neighbourhood_name', 'genus_name', 'tree_count']
).properties(
    width=500,
    height=300,
    title='Tree Count by Genus for Kensington-Cedar Cottage and Kitsilano'
)

scatter
```

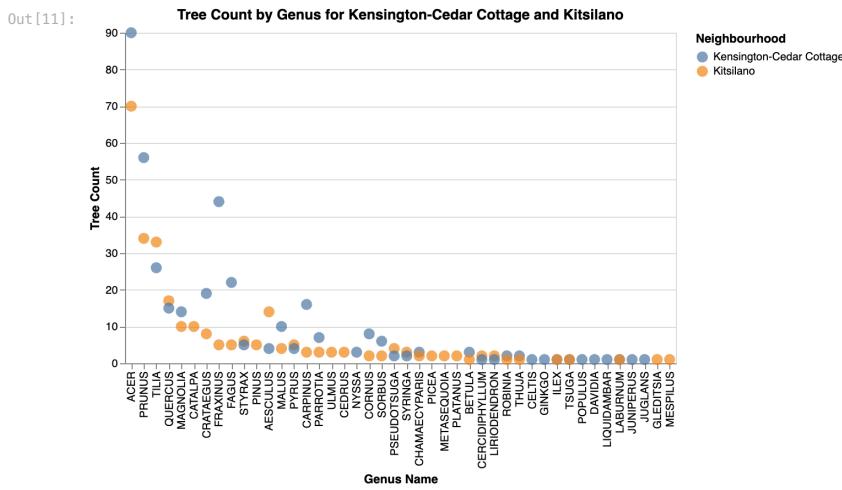


Figure 2 - Scatterplot of genus name and tree count in the Kensington-Cedar Cottage and Kitsilano neighbourhoods.

Figure 2 shows us that Fraxinus is the third most planted genus in Kensington-Cedar Cottage. Pinus in Kitsilano is observed with 5 trees under the 10 tree count line where most data points lie. Kensington-Cedar Cottage shows slightly less diversity and higher preference for some genera. It is evident in this chart that most genera are planted less than 5 times, especially in Kitsilano. This shows that the city of Vancouver does value diversity when it comes to planting choices. This likely results in increased ecosystem resilience, including increased wildlife support around the ecosystem, complex interactions resulting in varied responses to stress, as well as many ecological niches being filled that help decrease risk to the area.

The most popular genus for both Kensington-Cedar Cottage and Kitsilano is Acer, which is the Maple genus. The second most popular in Kensington-Cedar Cottage and Kitsilano is Prunus, which includes the popular Cherry Blossom. The third most popular genre for the Kensington-Cedar Cottage neighbourhood is the Fraxinus genus, the preferred genus for the emerald ash borer. Lastly, the third most popular genre for the Kitsilano neighbourhood is the Tilia. Overall, Figure 2 implies a high level of genera diversity in both neighbourhoods, despite some genera being preferred over others.

Having an understanding of the genera present in Kensington-Cedar Cottage and Kitsilano, it will also be helpful to investigate the diameter in these two neighbourhoods. This has high implications as in Kensington-Cedar Cottage, larger-diameter Fraxinus trees would be especially pricey to treat from an emerald borer infestation, and highly costly to remove as well. Their loss would also likely signify the loss of significant canopy which provides shade in urban areas and important ecosystem services. In relation to Kitsilano, a larger diameter would signify larger biomass which represents larger fuel loads that may worsen wildfires, especially in the context of the flammable Pinus genus.

```
In [12]: # Graph boxplot for diameter
boxplot = alt.Chart(top_neigh_df).mark_boxplot(extent='min-max').encode(
    x=alt.X('diameter:Q', title='Tree Diameter at Breast Height'),
    y=alt.Y('neighbourhood_name:N', title='Neighbourhood'),
    color=alt.Color('neighbourhood_name:N', title='Neighbourhood'),
    tooltip=['neighbourhood_name:N', 'diameter:Q']
).properties(
    title='Distribution of Tree Diameter in Kensington-Cedar Cottage and Kitsilano'
)

boxplot
```

Out [12]:

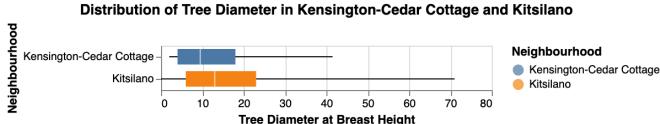


Figure 3 - Tree diameter at breast height (DBH) box plot across Kensington-Cedar Cottage neighbourhood and Kitsilano neighbourhood.

Figure 3 highlights that Kensington-Cedar Cottage possesses mostly moderate diameters. However, for some trees, diameter at breast height can get up to 41.5cm, which is quite a large tree. Although most trees don't fall into this category, some do, and this indicates that preventative action against the emerald ash borer is important in the Kensington-Cedar Cottage neighbourhood. This is especially true considering the high number of *Flaxinus* trees in Kensington-Cedar Cottage. Furthermore, although the emerald ash borer prefers trees in the *Flaxinus* genus as its host, once it has infected all *Flaxinus* trees, it would move on to trees belonging to other genera in the area, further highlighting the importance of taking preventative action. The species with higher diameter wouldn't necessarily have to be *Flaxinus* to be infected. Moreover, larger diameter trees provide crucial ecological services to the neighbourhood and would also be more difficult to save, and more costly to remove if infected. Kitsilano, on the other hand, possesses a tree diameter range larger than Kensington-Cedar Cottage, with values of up to 71cm. This value was observed in Table 2 as the highest diameter value present in the data frame. This indicates significant fuel loads if a wildfire occurs in the region, and emphasizes the need for wildfire safety measures to be put in place around the neighbourhood. Similarly to trees that aren't *Flaxinus* being at risk in the Kensington-Cedar Cottage neighbourhood, trees that aren't *Pinus* would also be at risk in the Kitsilano neighbourhood if a wildfire occurs. A wildfire burning a tree in the *Pinus* genus would be more flammable, but a wildfire in higher diameter trees would result in more fuel in the case of any genera, despite flammability varying. It is also important to note the potential of flammable trees in the *Pinus* tree spreading wildfire to other genera of large diameter.

Having an understanding of Kensington-Cedar Cottage's and Kitsilano's resilience and risk level, it will be interesting to investigate how they compare to the overall resilience and risk level across all neighbourhoods. This can be done by investigating biodiversity score across all neighbourhoods. By graphing total tree count and unique genera count, there can be an increased understanding of how much diversity there is in each neighbourhood relative to the total trees present.

```
In [13]: # Count trees per neighbourhood
tree_counts = van_trees['neighbourhood_name'].value_counts()

# Convert this to a dataframe
tree_counts = tree_counts.rename_axis('neighbourhood_name').reset_index(name='tree_count')

# Calculate number of genera
biodiversity = van_trees.groupby('neighbourhood_name')['genus_name'].nunique().reset_index(name='biodiversity_score')

# Merge
biodiversity = biodiversity.merge(tree_counts, on='neighbourhood_name')

# Create diversity graph for all neighbourhoods
bio_scatter = alt.Chart(biodiversity).mark_circle(size=100).encode(
    x=alt.X('tree_count:Q', title='Total Tree Count per Neighbourhood'),
    y=alt.Y('biodiversity_score:Q', title='Total Unique Genera per Neighbourhood'),
    color=alt.Color('biodiversity_score:Q', scale=alt.Scale(scheme='turbo'), title='Total Unique Genera'),
    tooltip=['neighbourhood_name', 'tree_count', 'biodiversity_score']
).properties(
    width=600,
    height=400,
    title='Total Tree Count and Unique Genera Count in Vancouver Neighbourhoods to Assess Biodiversity'
)

bio_scatter
```

Out [13]:

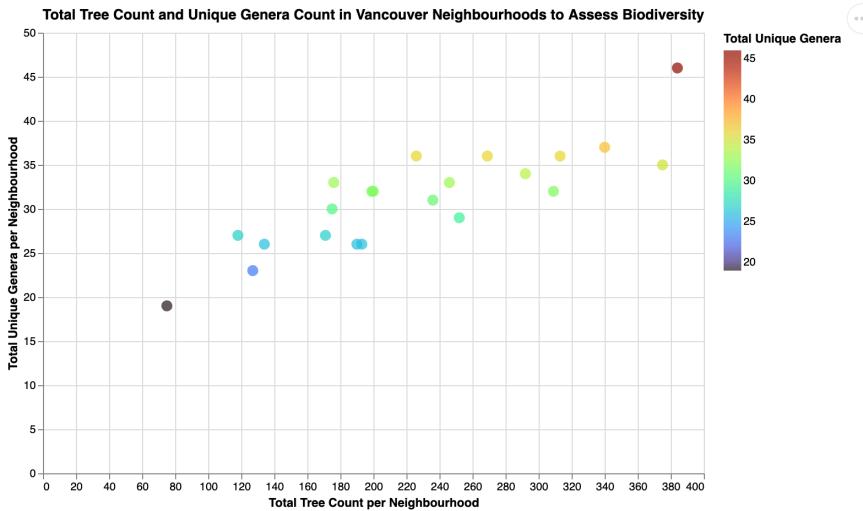


Figure 4 - Scatterplot of total tree count over unique genera count across all neighbourhoods.

Figure 4 displays what appears to be a positive relationship between total tree count and unique genera. This makes sense as the more trees are planted, the more variety in genera is possible. Kitsilano shows 269 total trees, and 36 genera. Kensington-Cedar Cottage, on the other hand, displays 375 total trees, and 35 genera. This implies that Kitsilano has higher biodiversity in comparison to Kensington-Cedar Cottage, which may be a result of how often Kensington-Cedar Cottage has planted its most popular genera (as seen in Figure 2). The overall trend shows that there is not too much diversity in terms of the relationship between tree count and unique genera. Most neighbourhoods fall within tree count of 118 and 375, and genera count of 23 and 37. There are two slight outliers, Strathcona with tree count of 75 and biodiversity score of 19, and Renfrew-Collingwood with tree count of 384 and biodiversity score of 46.

High diversity can increase ecosystem resilience and allow these species to be stronger in the fight against wildfires and the emerald ash borer. Moreover, varied genera is sometimes an approach taken by cities to protect urban ecosystems. Ideally, neighbourhoods would have higher genera diversity than what is observed here, however, the city may have had reasons to plant numerous trees of the same genera together in the same neighbourhoods, such as for aesthetic reasons.

An interesting way to further investigate risk level across neighbourhoods would be to graph the height\_range\_id category. It is known that if trees are similar in height, crown fires can occur more easily. Crown fires are considered some of the worst types of wildfire. This is the stage where the fire travels from each tree's crowns, burning through flammable foliage and easily sparking fires in trees located at the same level as the one already enlightened. For the height\_range\_id category, a category was given for trees depending on their height (0 = 0-10 ft, 1 = 10-20 ft, 2 = 20-30 ft, and 10 = 100+ ft). Let's graph this to see if trees tend to be closer or further from one another in terms of height categories. A small variety in height would indicate high crown fire susceptibility as sparks can easily spread from one crown to another.

```
In [14]: # Convert height_range_id to int
van_trees['height_range_id'] = van_trees['height_range_id'].astype(int)

# Group by neighbourhood_name and height_range_id
chart_data = van_trees.groupby(['neighbourhood_name', 'height_range_id']).size().reset_index(name='count')

# Obtain ordered list of height_range_id
height_order = sorted(chart_data['height_range_id'].unique())

# Sort neighbourhoods based on total tree count
neighbourhood_order = (
    chart_data.groupby('neighbourhood_name')['count']
    .sum()
    .sort_values(ascending=False)
    .index
    .tolist()
)

In [15]: # Create plot for height_range_id and neighbourhoods, sorted by neighbourhood_order
neighbourhood_order = (
    chart_data
    .groupby('neighbourhood_name')['count']
    .sum()
    .sort_values(ascending=False)
    .index
    .tolist()
)

height_neighbourhood = alt.Chart(chart_data).mark_circle().encode(
    x=alt.X('neighbourhood_name:N', title='Neighbourhood Name', sort=neighbourhood_order),
    y=alt.Y('height_range_id:N', title='Height Range ID Category'),
    size=alt.Size('count:Q', title='Tree Count'),
    color=alt.Color('count:Q', scale=alt.Scale(scheme='viridis')),
    tooltip=['neighbourhood_name:N', 'height_range_id:N', 'count:Q']
).properties(
    width=800,
    height=400,
    title='Tree Height Range ID Category by Neighbourhood'
).configure_axisX(labelAngle=-45)

height_neighbourhood
```

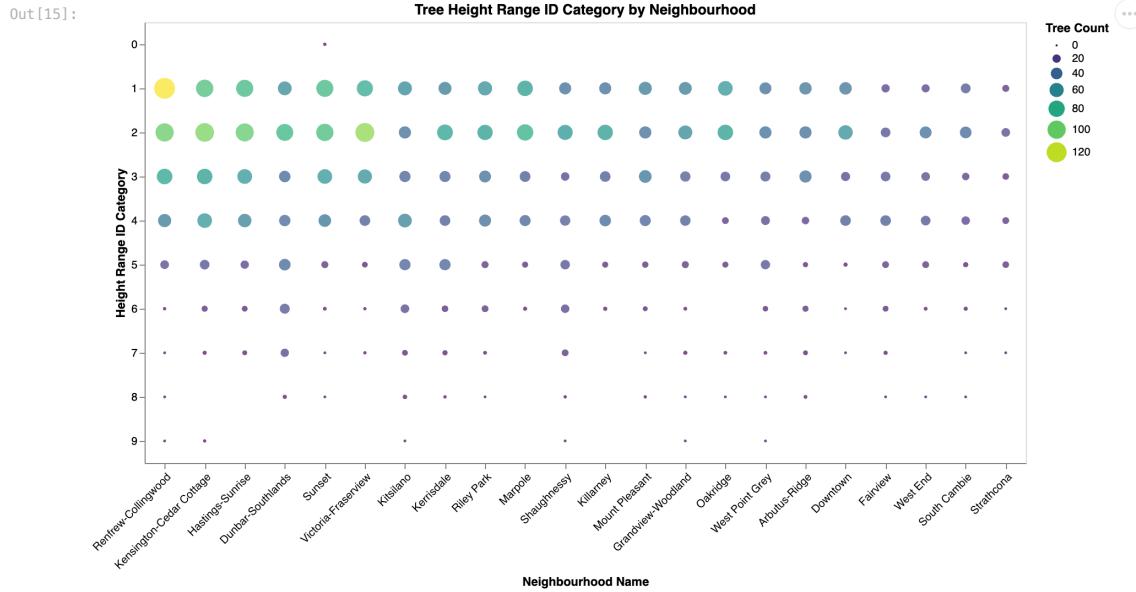


Figure 5 - Bubble plot of height\_range\_id (categorical height range value) and neighbourhood name colored by tree count.

As previously mentioned, the reference for height\_range\_id categories is: 0 = 0-10 ft, 1 = 10-20 ft, 2 = 20-30 ft, and 10 = 100+ ft. Figure 5 shows us that most of the trees tend to be approximately between categories 1-4, which signifies 10-50 feet. This is quite a wide range and it makes sense that most trees would fall within this range, as the remaining would either be less than 10 feet or over 50 feet tall. Under 10 feet, it is quite likely that many tree species would still be considered seedlings, and perhaps this is the reason they weren't included here. Further, it may also be that there are no new small trees around Vancouver based on the amount of time that passed between the last plantation, and this survey - most trees may have already grown tall enough, or may even been planted as a tall seedling rather than a seed.

Looking at each neighbourhood individually, it is evident that most neighbourhoods have a lot of trees that occur in the same height category. For instance, Renfrew-Collingwood has 132 trees of height category 1, Hastings-Sunrise has 88 under the same category, and Kensington-Cedar-Cottage has 91. Under category 2 we see a high quantity of trees as well in some neighbourhoods, with Hastings-Sunrise possessing 98, Kensington-Cedar-Cottage 105, Renfrew-Collingwood 100, and Victoria-Fraserview 108. It is important to note that a map showing the exact location of the trees is necessary to confirm that if a crown fire occurred, these trees would be in close enough proximity for this fire to travel between one another. Despite the trees' exact locations not being shown here, it is known that "fuel" layers are a highly important component of how a wildfire behaves, and Figure 3 makes it clear that there are possibly very strong fuel layers across numerous height categories in these neighbourhoods.

## Discussion

The visualizations presented here are highly helpful when re-visiting the questions proposed in the introduction.

1. **What neighbourhood around Vancouver is most susceptible to be negatively impacted by the emerald ash borer and uncontrollable (highly flammable) wildfires?** As mentioned earlier, the *Fraxinus* genus is the preferred host of the emerald ash borer, and the *Pinus* genus is known to be a highly flammable genus. Figure 1 described Kensington-Cedar Cottage as the neighbourhood with the most trees in the *Fraxinus* genus, with a total of 44. The second highest neighbourhood as shown in Figure 1 is Hastings-Sunrise, with 22 trees in the *Fraxinus* genus, half of what was observed in Kensington-Cedar Cottage. In relation to the *Pinus* genus, Kitsilano had the highest count of *Pinus* trees, although this total was only of five individuals, and may not impact wildfire risk in a significant manner in Kitsilano. The second highest was Strathcona with four trees of the genus *Pinus*. When interpreted by the genera presented here, the answer to the question would thus be, according to the data, Kensington-Cedar Cottage is the neighbourhood most likely to be negatively impacted by the emerald ash borer (with 44 *Fraxinus* trees), and Kitsilano is the neighbourhood with the most risk of highly flammable wildfires, although it is important to note that it only has five trees of the genus *Pinus* and a number this small might not pose a significant risk.
2. **What understanding do we have of these at-risk neighbourhoods (i.e. biodiversity, tree diameter...) that may help us infer how they might respond to these impacts?** Figure 2 showed high biodiversity in both Kensington-Cedar Cottage and Kitsilano. In Kitsilano, it was evident that trees in the same genus were often planted only less than five times, and there was high variety in genera. In Kensington-Cedar Cottage, the diversity level was similar but slightly less, as it was observed that there are some genera that are preferred by a high factor. As previously mentioned, preferred trees such as Maple and Cherry Blossom might be a result of aesthetic preferences in these spaces. Overall, Figure 2 showed that these neighbourhoods likely have increased risk against the negative impacts of climate change as a result of the spread amongst different genera in plantation choices. This helps increase overall ecosystem resilience against environmental impacts. On the other hand, when looking at diameter at breast height, Kensington-Cedar Cottage showing moderate diameter at breast height (Figure 3), with the highest at 41.5cm. Kitsilano showed high diameter at breast height in comparison to Kensington-Cedar Cottage, with values up to 71cm (Figure 3), the highest in the data frame (Table 2). The median for diameter at breast height for DBH is 9.5cm, and for Kitsilano, it is 13cm. Although the overall trend doesn't show high cause for concern, the highest values show that there are trees present in both neighbourhoods that would be highly problematic if infested by the emerald ash borer (Kensington-Cedar Cottage), or if a wildfire started (Kitsilano neighbourhood). Treating large trees infected by the emerald ash borer trees is highly costly, and removing dead ones would remove important shade, and crucial ecosystem services. Further, the flammable material that larger trees feed into a wildfire is crucial to understand, as this would cause significant impacts should a wildfire start in the region. With this in mind, it is important that the city evaluates its current plans, and considers providing extra protection against the emerald ash borer in large *Fraxinus* trees in Kensington-Cedar Cottage, and wildfire barriers where possible in *Pinus* trees in Kitsilano.
3. **How do these two at-risk neighbourhoods compare to all neighbourhoods in Vancouver in terms of biodiversity? Is their biodiversity higher in a way that might imply increased ecosystem resilience against invasive species and wildfire?** Figure 4 provided a relative idea of biodiversity levels in the different neighbourhoods. Kitsilano displayed 269 total trees, and 36 genera. Kensington-Cedar Cottage, on the other hand, displayed 375 total trees, and 35 genera. From the figure, it can be implied that Kitsilano has slightly higher biodiversity in comparison to Kensington-Cedar Cottage. This further corroborates what was observed when analyzing Figure 2. Figure 4 shows that there is not much variety between the different neighbourhoods and their relationship between unique genera and tree count. To answer the question presented here, the biodiversity level in Kitsilano and Kensington-Cedar Cottage do not imply a difference in level between the rest of the neighbourhoods. Most neighbourhoods were observed within the same general area of tree count (118-375) and genera (23-37). Strathcona (count of 75, genera total of 19) and Renfrew-Collingwood (count of 384, genera total of 46) were slight outliers in the data.
4. **What might tree height across all different neighbourhoods show about potential crown fire spread (one of the most severe types of wildfire - where wildfire travels from one tree top to another) in Vancouver?** Figure 5 showed a concerning number of trees in the same height\_id\_range category. As previously mentioned, a map of these is needed in order to know the exact distance between the trees, and if similar heights would imply an increased potential of crown fire spread or not. However, Figure 5 does show that, even if only some of these trees are close together, they could increase the risk of crown fire in these neighbourhoods, as a result of how close their heights are to one another. This is evident as under height\_id\_range category 1, Renfrew-Collingwood has 132 trees, Hastings-Sunrise has 88, and Kensington-Cedar-Cottage has 91. Under category 2, a high value pattern can be observed as well, with Hastings-Sunrise possessing 98, Kensington-Cedar-Cottage 105, Renfrew-Collingwood 100, and Victoria-Fraserview 108. It is important for the city to be aware of the high risk that crown fires pose, especially in urban areas near homes, and in the context of the significant increases in wildfire occurrences in BC in the last decade.

Use of graph type, colors, and shape all played a large role in enabling the visualizations to present the data with clarity. For Figure 1, I believe a bar chart was a great choice to see how the two different genera at hand were distributed between neighbourhoods. The color choices made the differences quite clear as well. Figures 2 and 3 also showed the data in an effective manner. Matching the colors between the two made it easy to see that they showed the same neighbourhoods, and the box plot choice worked well for the diameter at breast height data as it provided valuable information on the spread of diameter at breast height. In relation to Figure 4, I believe the color choice here was highly useful in displaying which neighbourhoods had the highest values of total genera. A scatterplot was a good way to portray the relationship between total tree count and genera count. Lastly, the use of the viridis color scheme for Figure 5 made tree count differences very easily interpreted, and displaying the data in this graph format made it very clear to see what was occurring in each neighbourhood. Furthermore, the angle of the neighbourhoods on the X-axis made it easier to read the neighbourhood names.

## Conclusion

In conclusion, neighbourhoods around Vancouver display a potentially high risk factor through the *Fraxinus* trees planted (susceptible to emerald ash borer), the *Pinus* trees planted (highly flammable if a wildfire occurs), the large tree diameter observed in some cases, as well as highly similar tree heights and the role this plays in increasing potential crown fire spread. However, biodiversity levels did not appear extremely concerning and this may be a factor that is increasing the resilience of these neighbourhoods' trees in the fight against climate change's growing risks. Further investigation is required in order to better understand what might be occurring, and how these areas can be further protected.

As previously mentioned, an interesting project to build off of this would be mapping these trees in the Vancouver area, or assessing their proximity to one another. This would help further simulate emerald ash borer infestation, wildfire spread, as well as how close together trees of different genera are planted (for biodiversity assessments), and how trees of different diameters and heights are spread around these areas. A map would enable the question presented here to be answered to a fuller extent, and with those answers, for further questions to be investigated. It would also be very interesting to use wildfire or invasive species monitoring tools to combine with this research to see how it may complement the current findings here. Furthermore, I am also curious about the animal species in the area and how these differ based on the genera types, and how these impact climate change risk and resilience levels. Although this project allowed for an understanding of some of these factors and trends, given ecosystems' complexities, different data would be needed from numerous different approaches to tell the full story.

Latly, the results presented here were not what I expected. I was surprised to see how many trees overlapped in height category in different neighbourhoods, as well as how in some cases, neighbourhoods showed a large bias towards a specific genus of trees. This intrigued me and made me curious about the selection process as to what gets planted. It is also interesting to consider that some of these trees are very old, and were planted before the city of Vancouver had its current system of monitoring in place. It would be interesting to investigate how the history of Vancouver played a role in the genera found around the city.

## Dashboard

```
In [16]: # Define parameters and selection
diameter_slider = alt.param(
    name='min_diameter',
    value=0,
    bind=alt.binding_range(min=0, max=71, step=1, name='Minimum Diameter Scale ')
)
```

```

genus_selection = alt.selection_point(fields=['genus_name'])

# Add "All" to options
neigh_dropdown = alt.binding_select(
    options=['All', 'Kensington-Cedar Cottage', 'Kitsilano'],
    name='Select Neighbourhood:'
)
neigh_selection = alt.param(name='neigh_param', value='All', bind=neigh_dropdown)

# Scatter plot with "all" in dropdown
scatter = alt.Chart(grouped).mark_circle(size=100).encode(
    x=alt.X('genus_name:N', title='Genus Name', sort='y'),
    y=alt.Y('tree_count:Q', title='Tree Count'),
    color=alt.Color('neighbourhood_name:N', title='Neighbourhood'),
    tooltip=[('neighbourhood_name', 'genus_name', 'tree_count')],
    opacity=alt.condition(genus_selection, alt.value(1), alt.value(0.3))
).add_params(
    genus_selection,
    neigh_selection,
    diameter_slider
).transform_filter(
    (alt.datum.neighbourhood_name == neigh_selection) | (neigh_selection == 'All')
).properties(
    width=500,
    height=300,
    title='Tree Count by Genus for Kensington-Cedar Cottage and Kitsilano'
)

# Setting fixed y-axis for boxplot so that it doesn't move based on difference in character length
# between neighbourhoods; creating boxplot
y_domain = ['Kensington-Cedar Cottage', 'Kitsilano']

boxplot = alt.Chart(top_neigh_df).mark_boxplot(extent='min-max').encode(
    x=alt.X('diameter:Q', title='Tree Diameter at Breast Height'),
    y=alt.Y('neighbourhood_name:N',
            title='Neighbourhood',
            scale=alt.Scale(domain=y_domain)),
    color=alt.Color('neighbourhood_name:N', title='Neighbourhood'),
    tooltip=[('neighbourhood_name:N', 'diameter:Q')]
).transform_filter(
    genus_selection
).transform_filter(
    (alt.datum.neighbourhood_name == neigh_selection) | (neigh_selection == 'All')
).transform_filter(
    'datum.diameter >= min_diameter'
).properties(
    width=500,
    height=300,
    title='Distribution of Tree Diameter in Kensington-Cedar Cottage and Kitsilano'
)

# Dashboard
dashboard = scatter | boxplot
dashboard

```

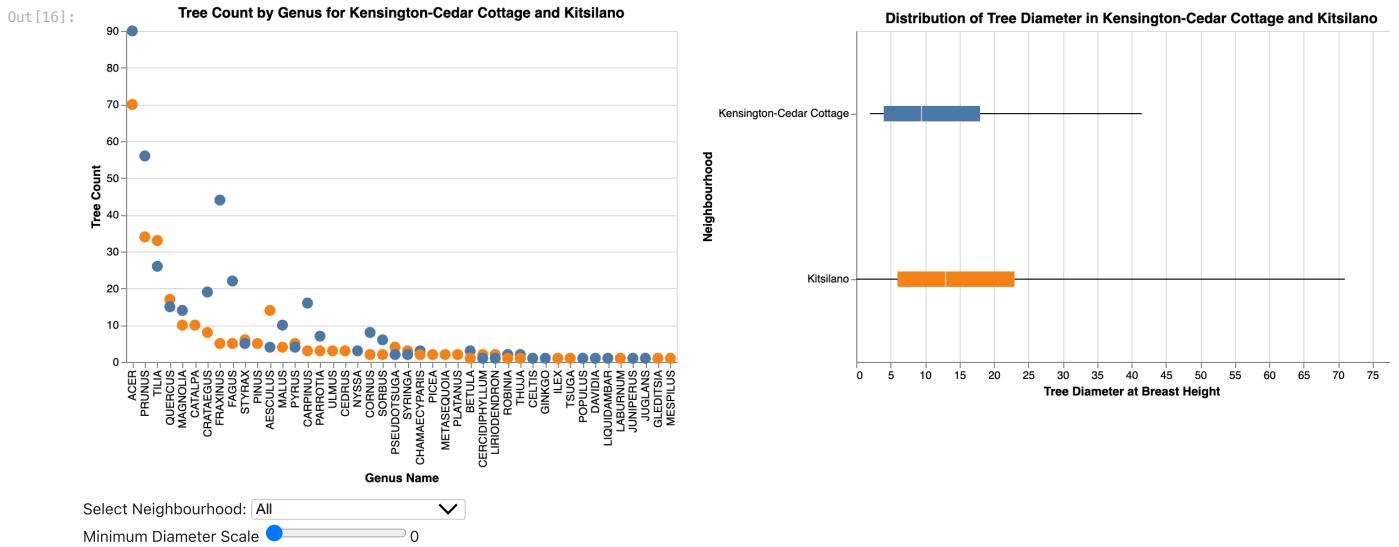


Figure 6 - Dashboard merging tree count scatterplot and tree diameter boxplot.

Using the dashboard provides a more detailed view of the relationships between the variables present. For instance, clicking on the scatterplot allows one to select any genus and observe its diameter breakdown in the box plot. Furthermore, the dropdown allows neighbourhood selection so that these variables can be analyzed individually between neighbourhoods if that is of interest. It is also helpful to have the tree count variable so that it can be compared to the diameter breakdown, and individual tree diameters can be compared to a group of tree diameters (depending on how many trees are in the genus selected). The diameter scale slider allows for a more detailed view of how diameter changes between neighbourhoods and any individual genus selected. For instance, one can select the Fraxinus genus and use the slider to observe its diameter spread across neighbourhoods, or select one of the two neighbourhoods using the dropdown menu in order to isolate the observations to just one area.

The design choices made here worked to ensure that numerous different components could be analysed in detail (by neighbourhood, by genera, by specific diameter size, amongst others). I wanted to focus on the two neighbourhoods selected at the beginning of this analysis, as I thought this breakdown in variety of different genera was quite interesting, as the neighbourhoods are similar to one another, yet distinct differences can be observed. Coupling this with the differences observed in the box plot further peaked my curiosity, as the neighbourhoods selected showed high variety in diameter at breast height, and I was aware that different genera would likely display differences in diameter at breast height as

well. This interactive dashboard allows these components to be analyzed further. If I had more time, I would try to place a widget under each graph, as I was unable to do this for this project, but I believe it would make the dashboard more visually appealing. I believe it would also have been interesting to include the other graphs here, and to tell the full story described in the discussion in the dashboard as well, rather than focusing solely on the Kensington-Cedar Cottage and Kitsilano neighbourhoods.

## References

- [Open Data Portal](#)
- [Data cleaned by course instructors](#)
- [Label Angle Altair Tutorial](#)
- [Wildfire management in Canada: Review, challenges and opportunities](#)
- [Flammability features of native and non-native woody species from the southernmost ecosystems: a review](#)
- [The relationship between the emerald ash borer \(\*Agrilus planipennis\*\) and ash \(\*Fraxinus spp.\*\) tree decline: Using visual canopy condition assessments and leaf isotope measurements to assess pest damage](#)
- [Experimental Study on Effects of Moisture Content and Tree Height on Crown Fire Behaviors of Live Cypress Trees](#)
- [Variables that influence changes in fire severity and their relationship with changes between surface and crown fires in a wind-driven wildfire](#)
- [Five Potential Consequences of Climate Change for Invasive Species](#)
- [Climate change increases the risk of wildfires](#)
- [Evaluating published approaches for modelling diameter at breast height from stump dimensions](#)
- [Avoiding getting burned: becoming FireSmart](#)
- [Altair Filter Transform](#)
- [Altair Datum](#)
- [Fuel layer specific pollutant emission factors for fire prone forest ecosystems of the western U.S. and Canada](#)