vencouver-trees

April 12, 2023

1 Vancouver street trees

1.1 Final Project Data Analysis

1.2 Fatemeh Salim

1.2.1 Motivation

As a resident of beautiful Vancouver, I truly believe part of its beauty is because of its trees, especially cherry trees that when bloom creates beautiful scenery. Trees also clean the air, absorbs rainwater, and provides bird habitat. I find it interesting to know which Vancouver neighbourhood has the greatest number of trees. which trees being planted most often in any of these neighbourhoods?

When it is cherry blossom blooming season, in which neighbourhood they can be found the most? Which neighbourhood has more tallest cherry trees? Different type of cherry trees may bloom in different times of the year. It would be useful to be able to investigate neighbourhoods for a specific kind of cherry tree. Here I am going to explore Vancouver trees dataset and answering following question.

1.2.2 Questions of interest

1. Which Vancouver neighbourhood has the greatest number of trees?
2. Which trees are most planted in each neighbourhood over the years?
3. Where are the most cherry trees in Vancouver located?
4. How height and diameter of trees in Vancouver related?

1.3 Analysis

1.3.1 Data Imports

For this project, I will be using a subset of the Vancouver Street Trees that can be found on City of Vancouver website.

With Altair it is not easy to locate Vancouver on the global map and there is no projection for Canada like there is for the United states, I used the geojson for Vancouver available through a URL that is obtained from the Vancouver Data Portal.

[ ]:
```
import altair as alt
import pandas as pd
alt.data_transformers.enable('default', max_rows=1000000)
```

```
import json
```

[ ]:
```
trees_df = pd.read_csv(
    "https://raw.githubusercontent.com/UBC-MDS/data_viz_wrangled/main/data/
 ↪Trees_data_sets/small_vancouver_trees.csv",
    parse_dates=["date_planted"],
)
```

[ ]:
```
trees_df.head()
```

[ ]:
```
   Unnamed: 0        std_street         on_street  species_name  \
0       19886         W 10TH AV         W 10TH AV  BIGNONIOIDES
1        7941         W 59TH AV         W 59TH AV   SACCHARINUM
2        4613         W 47TH AV         W 47TH AV   PLATANOIDES
3        7388  COMMERCIAL DRIVE  COMMERCIAL DRIVE  EUCHLORA   X
4        1894         E 55TH AV         E 55TH AV       SPECIES

    neighbourhood_name date_planted  diameter street_side_name genus_name  \
0            Kitsilano          NaT      34.0              ODD    CATALPA
1              Marpole          NaT      20.0              ODD       ACER
2            Kerrisdale          NaT      24.0              ODD       ACER
3   Grandview-Woodland          NaT       8.0             EVEN      TILIA
4  Victoria-Fraserview          NaT      14.0             EVEN      ABIES

  assigned  …  plant_area curb tree_id                common_name  \
0        N  …          10    Y    9945               COMMON CATALPA
1        Y  …          16    Y   50427                 SILVER MAPLE
2        N  …          12    Y   43456                 NORWAY MAPLE
3        N  …           C    Y   69099               CRIMEAN LINDEN
4        N  …           B    Y  164752  CRIMSON SUNSET NORWAY MAPLE

  height_range_id  on_street_block  cultivar_name root_barrier   latitude  \
0               5             3200            NaN            N  49.263400
1               4              700            NaN            N  49.217059
2               5             2200            NaN            N  49.229119
3               3             1300            NaN            N  49.272647
4               5             1900            NaN            N  49.219958

    longitude
0 -123.177100
1 -123.120787
2 -123.159841
3 -123.069463
4 -123.067159

[5 rows x 21 columns]
```

1.3.2 Dataset description

The below descriptions are from this website where the dataset was obtained.

"The street tree dataset includes a listing of public trees on boulevards in the City of Vancouver and provides data on tree coordinates, species and other related characteristics. Park trees and private trees are not included in the inventory." This table contains different information about tree common name, neighbourhood, date planted, height range, diameter, species name, genus name, and more.

Here is a brief description of the columns of this table:

Column	Description
Numerical ID	identifier
CIVIC_NUMBER	Street address of the site at which the tree is associated with
STD_STREET	Street name of the site at which the tree is associated with
GENUS_NAME	Genus's name
SPECIES_NAME	Species name
CULTIVAR_NAME	Cultivar name
Common name	Name of tree
ASSIGNED	Indicates whether the address is made up to associate the tree with a nearby lot (Y=Yes or N=No)
ROOT_BARRIER	Root barrier installed (Y = Yes, N = No)
PLANT_AREA	B = behind sidewalk, G = in tree grate, N = no sidewalk, C = cutout, a number indicates boulevard width in feet
ON_STREET_BLOCK	The street block at which the tree is physically located on
ON_STREET	The name of the street at which the tree is physically located on
NEIGHBOURHOOD_NAME	City's defined local area in which the tree is located
STREET_SIDE_NAME	The street side which the tree is physically located on (Even, Odd or Median (Med))
HEIGHT_RANGE_ID	0-10 for every 10 feet (e.g., 0 = 0-10 ft, 1 = 10-20 ft, 2 = 20-30 ft, and 10 = 100+ ft)
DIAMETER	DBH in inches (DBH stands for diameter of tree at breast height)
CURB	Curb presence (Y = Yes, N = No)
DATE_PLANTED	The date of planting in YYYYMMDD format. Data for this field may not be available for all trees.

Before advancing any further, lets explore the data set first and pick the columns that will be used in answering my questions.

[ ]:
```
trees_df.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         2338 non-null   datetime64[ns]
 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           4963 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        2700 non-null   object
 18  root_barrier         5000 non-null   object
 19  latitude             5000 non-null   float64
 20  longitude            5000 non-null   float64
dtypes: datetime64[ns](1), float64(3), int64(5), object(12)
memory usage: 820.4+ KB
```

date_planted has about half of its data missing. Although this data could add very interesting layer to my analysis, but I decided to exclude this column. For answering my question, I will be using the following columns only:

[ ]:
```
trees_df = trees_df[
    [
        "neighbourhood_name",
        "diameter",
        "common_name",
        "height_range_id",
        "latitude",
        "longitude",
    ]
]
trees_df
trees_df = trees_df.rename(columns={"neighbourhood_name": "name"})
```

[ ]:
```
trees_df.describe(exclude="number", datetime_is_numeric=True)
```

[ ]:

	name	common_name
count	5000	5000
unique	22	339
top	Kensington-Cedar Cottage	KWANZAN FLOWERING CHERRY
freq	441	363

[ ]:
```
trees_df.describe()
```

[ ]:

	diameter	height_range_id	latitude	longitude
count	5000.000000	5000.000000	5000.000000	5000.000000
mean	12.132900	2.699800	49.247739	-123.105449
std	9.310923	1.550923	0.020973	0.049506
min	0.250000	0.000000	49.201366	-123.223440
25%	4.250000	2.000000	49.230902	-123.144000
50%	10.000000	2.000000	49.248583	-123.102044
75%	17.000000	4.000000	49.263816	-123.062371
max	182.000000	9.000000	49.293881	-123.022469

2 Question 1: Which Vancouver neighbourhoods has the most number of trees?

Let's start with the map of Vancouver. It will be easier to locate neighbourhoods on the map.

[ ]:
```
url_geojson = 'https://raw.githubusercontent.com/UBC-MDS/exploratory-data-viz/
↪main/data/local-area-boundary.geojson'
```

[ ]:
```
data_geojson_remote = alt.Data(url=url_geojson, format=alt.
↪DataFormat(property='features',type='json'))

data_geojson_remote
```

[ ]:
```
Data({
  format: DataFormat({
    property: 'features',
    type: 'json'
  }),
  url: 'https://raw.githubusercontent.com/UBC-MDS/exploratory-data-
viz/main/data/local-area-boundary.geojson'
})
```

[ ]:
```
vancouver_map = alt.Chart(data_geojson_remote).mark_geoshape(
    color = 'gray', opacity= 0.5, stroke='white').encode(
).project(type='identity', reflectY=True)
#vancouver_map
```

[ ]:

```
count_df = trees_df.groupby("name")["name"].count().
 ↪reset_index(name='tree_count')
count_df

points_df = trees_df.groupby("name")["longitude",'latitude'].median()#.
 ↪reset_index()
points_df

counts_df = count_df.merge(points_df, on ="name")
#counts_df
```

```
C:\Users\fatem\AppData\Local\Temp\ipykernel_6296\522321266.py:4: FutureWarning:
Indexing with multiple keys (implicitly converted to a tuple of keys) will be
deprecated, use a list instead.
  points_df =
trees_df.groupby("name")["longitude",'latitude'].median()#.reset_index()
```

[ ]:
```
points = (
    alt.Chart(counts_df)
    .mark_circle()
    .encode(
        longitude="longitude",
        latitude="latitude",
        size="tree_count:Q",
        color=alt.Color("tree_count:Q", title="Tree count"),
        tooltip=["name:N", alt.Tooltip("tree_count:Q", title="Tree counts")],
    )
    .project(type="identity", reflectY=True)
    .properties(height=300, width=600, title="Vancouver neighbourhoods")
)
van_map_points = vancouver_map + points
van_map_points
```

```
c:\Users\fatem\AppData\Local\Programs\Python\Python39\lib\site-
packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and
will be removed in a future version. Use .items instead.
  for col_name, dtype in df.dtypes.iteritems():
```

[ ]:
```
alt.LayerChart(…)
```

I am going to try choropleth map as well and will decide which map is more helpful here.

[ ]:
```
title = alt.TitleParams(
    "Kensington-Cedar Cottage has the most number of trees",
    subtitle="Neighbourhoods are clickable",
)

van_map = (
```

```
    alt.Chart(data_geojson_remote)
    .mark_geoshape()
    .transform_lookup(
        lookup="properties.name",
        from_=alt.LookupData(counts_df, "name", ["tree_count", "name"]),
    )
    .encode(
        color=alt.Color("tree_count:Q", title=" Tree count"),
        tooltip=["name:N", alt.Tooltip("tree_count:Q", title="Tree counts")],
    )
    .project(type="identity", reflectY=True)
    .properties(title=title)
)
van_map


# Add Labels Layer
labels = (
    alt.Chart(counts_df)
    .mark_text()
    .encode(
        longitude="longitude",
        latitude="latitude",
        text="name:N",
        size=alt.value(8),
        opacity=alt.value(1),
    )
    .project(type="identity", reflectY=True)
    .properties(height=300, width=600, title="Vancouver map")
)

van_map = van_map + labels
van_map
```

[ ]:
```
alt.LayerChart(…)
```

I will continue with choropleth map, since it is easier to distinguish counts of trees by color in this map.

We can tell from the above map that **Kensington-Cedar Cottage**, **Renfrew-Collingwood**, and **Hastings-Sunrise** with 441, 404, and 371 trees respectively are the top three neighbourhoods in terms of number of trees planted.

Strathacona with only 91 trees had the least number of trees.

Now that we know neighbourhoods' tree count ,the next question will be about the most popular trees in each of these neighbourhood.

3 Question 2: Which trees are mostly planted in each neighbourhood over the years?

How I would like to answer this question is by fisrt accessing each neighbourhood/neighbourhoods through the map.

[ ]:
```
click = alt.selection_multi(fields=["name"])

van_map_click = van_map.encode(
    opacity=alt.condition(click, alt.value(1), alt.value(0.3))
).add_selection(click)
```

[ ]:
```
top_popular_trees = (
    alt.Chart(trees_df)
    .transform_filter(click)  # filter for selected neighbourhood
    .mark_bar()
    .encode(
        alt.X("count():Q", title=""),
        alt.Y("common_name:N", title="", sort="x"),
        color="height_range_id:N",
        tooltip=[alt.Tooltip("count():Q", title="")],
    )
)
```

[ ]:
```
# Adding slider to contol the number of top popular trees being shown on bar␣
 ↪chart

slider = alt.binding_range(
    name="Select the number of top popular trees you want to see: ",
    step=1,
    min=5,
    max=25)

select_trees = alt.selection_single(
    fields=["num_names"], init={"num_names": 20}, bind = slider)
```

[ ]:
```
title = alt.TitleParams(
    "Most popular trees in selected neighbourhood(s)",
    subtitle="Kwanzan Flowering Cherry tree is very popular",
)
top_names = (
    alt.Chart(trees_df)
    .transform_filter(click)  # filter for selected neighbourhood
    .mark_bar()
    .encode(
        alt.X("count:Q", title=""),
        alt.Y("common_name:N", title="", sort="-x"),
```

```
    )
    .transform_aggregate(count="count()", groupby=["common_name"])
    .transform_window(
        rank="rank(count)", sort=[alt.SortField("count", order="descending")]
    )
    .transform_filter(alt.datum.rank <= select_trees.num_names)
    .properties(title=title, height=400, width=300)
    .add_selection(click)
    .add_selection(select_trees)
)

van_map_click | top_names
```

[ ]:
```
alt.HConcatChart(…)
```

When all neighbourhoods are selected on the map, we can see that **Kwanzan flowering Cherry**, **Pissard plum**, and **Norway maple** are the top tree popular trees in whole Vancouver.

We can click on each neighbourhood and quuickly discover that **Kwanzan flowering cherry** trees always appears as one of the most popular trees in every individual neighbourhood, except downtown. So, let's explore Kwanzan flowering cherry as well as other cherry trees in more depth in the next question.

4 Question 3: Where are the most cherry trees in Vancouver located?

[ ]:
```
cherry_trees = trees_df[trees_df["common_name"].str.contains("CHERRY")]

# finding most popular cherry trees in vancouver
top_cherry_trees = (
    cherry_trees.groupby("common_name")["common_name"]
    .count()
    .reset_index(name="count")
    .sort_values(by="count", ascending=False).iloc[:6,0].tolist()
)
cherry_trees = cherry_trees [cherry_trees["common_name"].isin(␣
 ↪top_cherry_trees)]
cherry_trees
```

[ ]:
```
                          name  diameter              common_name  \
6                    West End      24.0  KWANZAN FLOWERING CHERRY
14         Victoria-Fraserview      16.0  KWANZAN FLOWERING CHERRY
19                     Marpole      15.0  AKEBONO FLOWERING CHERRY
23              Mount Pleasant      26.0    PINK PERFECTION CHERRY
27           Grandview-Woodland       9.0     RANCHO SARGENT CHERRY
…                           …       …                       …
4928  Kensington-Cedar Cottage      24.5  KWANZAN FLOWERING CHERRY
```

```
4962                  Oakridge      19.5  KWANZAN FLOWERING CHERRY
4976        Grandview-Woodland      29.0  KWANZAN FLOWERING CHERRY
4981             Arbutus-Ridge      10.0  KWANZAN FLOWERING CHERRY
4987       Victoria-Fraserview      12.0  KWANZAN FLOWERING CHERRY

      height_range_id   latitude   longitude
6                   3  49.286839 -123.131659
14                  3  49.218128 -123.070469
19                  2  49.212336 -123.115185
23                  4  49.265306 -123.091927
27                  3  49.270114 -123.065648
…                 …        …          …
4928                2  49.251731 -123.074946
4962                2  49.228831 -123.113102
4976                3  49.275683 -123.066599
4981                2  49.254542 -123.166197
4987                3  49.218388 -123.073899

[522 rows x 6 columns]
```

[ ]:
```
title = alt.TitleParams(
                "Cherry trees in neighbourhood(s) , clickable",
    subtitle=[ "Mount Pleasent has the most number of cherry trees","downtown␣
 ↪vancouver has the least"],
)


sort_order = [1, 2, 3, 4]
neighbourhood_cherry = (
    alt.Chart(cherry_trees, title=title)
    .mark_bar()
    .encode(
        alt.X("count()"),
        alt.Y("name", sort=sort_order, title=""),
        color=alt.Color("common_name:N", title = "Cherry trees"),
        opacity=alt.condition(click, alt.value(1), alt.value(0.2)),
    )
    .add_selection(click)
    .properties(height=400, width=300)
)
(van_map_click | neighbourhood_cherry)
```

```
c:\Users\fatem\AppData\Local\Programs\Python\Python39\lib\site-
packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and
will be removed in a future version. Use .items instead.
  for col_name, dtype in df.dtypes.iteritems():
```

[ ]:
```
alt.HConcatChart(…)
```

Mount pleasant must be beautiful in spring. It has the greatest number of cherry trees and majority of them are of type **Kwanzan flowerring cherry**.

Downton Vancouver has just less than 5 cherry trees.

There are different kinds of cherry which means we have flowers from February to June. **Akebono** and **Kwanzan** are very popular. Akebono blooms first, Kwanzan is a week or two after that.

It would be great to be able to narrow down to tree(s) of interest based on the time of the year we plan to visit them. Let's make the legend in above chart clickable to be able to explore different kinds of cherry trees more.

[ ]:
```
click_legend = alt.selection_multi(fields=['common_name'], bind='legend')
title = alt.TitleParams(
    "Mount Pleasent neighbourhood has the most number of cherry trees",
    subtitle="downtown vancouver has least cherry trees",
)

sort_order = [1, 2, 3, 4]

# Multiple selections from legend
neighbourhood_cherry_base = (
    alt.Chart(cherry_trees, title=title)
    .mark_bar()
    .encode(
        alt.X("count()"),
        alt.Y("name", sort=sort_order, title="Neighbourhood"),
        color=alt.Color("common_name:N", title = "Click on cherry tree(s) of␣
 ↪intrest")#,
        #opacity=alt.condition(click, alt.value(1), alt.value(0.2))
    )
    #.add_selection(click)
    .properties(height=400, width=300)
)

background = neighbourhood_cherry_base .mark_bar(opacity=0)
forground= neighbourhood_cherry_base.add_selection(click_legend).
 ↪transform_filter(click_legend)

neighbourhood_cherry_base = background + forground

neighbourhood_cherry_base
#(van_map_click | neighbourhood_cherry).add_selection(click_legend)????
```

[ ]:
```
alt.LayerChart(…)
```

5 Question 4: How height and diameter of trees in Vancouver related?

To answer this question, I will take a look at top 25 popular trees. Tree common name can be selected from dropdown.

[ ]:
```
common_trees = (
    trees_df["common_name"]
    .value_counts()[:25]
    .sort_values(ascending=False)
    .reset_index(name="count")
)
common_trees

tree_names = sorted(common_trees["index"].unique())
dropdown = alt.binding_select(
    name="Select one of the top popular trees in Vancouver to see height and␣
 ↪diameter relationship   ",
    options=tree_names,
)
select_tree = alt.selection_single(fields=["common_name"], bind=dropdown)
```

[ ]:
```
tree_size_plot_scatter = (
    alt.Chart(trees_df[trees_df["diameter"] < 80])
    .mark_circle()
    .encode(alt.X("diameter", title="Diameter (inch)"), alt.
 ↪Y("height_range_id"))
).transform_filter(select_tree)

tree_size_plot_line = (
    alt.Chart(trees_df)
    .mark_line(color="Red")
    .encode(
        alt.X("mean(diameter)"),
        alt.Y("height_range_id", title="Height range Id"),
        tooltip=alt.value("Mean of diameter"),
    ).properties(height = 250, width = 770, title = "Relationship between␣
 ↪height and diamter of popular trees in Vancouver")
).transform_filter(select_tree)
tree_size = tree_size_plot_line + tree_size_plot_scatter

# van_map_click |(tree_size_plot_line + tree_size_plot_scatter).
 ↪add_selection(click)

tree_size = tree_size.add_selection( click).add_selection(click).
 ↪add_selection(select_tree).transform_filter(select_tree)
tree_size
```

```
c:\Users\fatem\AppData\Local\Programs\Python\Python39\lib\site-
packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and
will be removed in a future version. Use .items instead.
  for col_name, dtype in df.dtypes.iteritems():
```

[ ]:
```
alt.LayerChart(…)
```

As we can tell from the above chart, there is a positive relation ship between the height and diameter of each of the popular trees in Vancouver.

However, we can tell it is not always the case that taller trees be thicker.

Also, we can tell from this chart that **Norway maple** trees can grow as tall as 90 ft.

6 Discussion

Vancouver trees has a significant importance since they add to the beauty of the city as well as they clean the air, absorb rainwater, and provide bird habitat. In my analysis I explored different neighbourhood of Vancouver first to see which one has the most trees in total.

As it turns out **Kensington-Cedar Cottage**, **Renfrew-Collingwood**, and **Hastings-Sunrise** with 441, 404, and 371 trees respectively are the top three neighbourhoods in terms of count of trees planted. **Strathacona** with only 91 trees had the least number of trees.

After this a question that stands out is what the most popular trees are in Vancouver as well as in every individual neighbourhood.

When all neighbourhoods are selected on the map, we can see that **Kwanzan flowering cherry**, **Pissard plum**, and **Norway maple** are the top three popular trees in whole Vancouver.

Also, we quickly discover that **Kwanzan flowering cherry** tress always appears as one of the most popular trees in every individual neighbourhood, except downtown, so it is very popular.

In fact, as spring nears, Vancouverites and tourists looking forward to cherry blossom that blanket streets and parks throughout the city so it worth knowing where the most of them are located.

I figured that **Mount pleasant** has the greatest number of cherry trees and majority of them are of type Kwanzan flowering cherry.

Downton Vancouver instead has just less than 5 cherry trees and is not a good candidate for visiting cherry trees during spring.

Different kinds of cherry trees bloom at different times of the year. The Legend of the cherry trees plot can be used to narrow down to specific kind of cherry and see their abundance in different neighbourhood(s).

Finally, we can see that popular trees in Vancouver that are taller in general has larger diameter. From the last plot we can tell how tall different trees can grow to. For example **Norway maple** trees can grow as tall as 90 ft.

This has been a very interesting dive into the Vancouver trees! In future, I would like to examine trend over year for popular trees in Vancouver and also how tree's age affects their height and diameter.

7 Dashboard

[ ]:
```
alt.themes.enable('none');
(
    van_map_click.properties(width = 750)
    & (top_names | neighbourhood_cherry).add_selection(click)
    & tree_size.add_selection(select_tree).transform_filter(select_tree))
# .configure_view(stroke=None)
```

[ ]:
```
alt.VConcatChart(…)
```

8 Reference

[website] https://opendata.vancouver.ca/explore/dataset/street-trees/information/?disjunctive.species_name&dis

news.ubc.ca