In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import json, ast
import plotly.express as px

sns.set()
```

In [2]:
```
import ipywidgets as widgets
from IPython.display import display
from IPython.display import Image
```

In [3]:
```
import plotly
import plotly.express as px
print(plotly.__version__)
```

```
5.1.0
```

In [4]:
```
from matplotlib.colors import LogNorm, Normalize
```

Research Questions

1. What are the factors facilitating the high tree density in the Allegheny Center, Friendship, and Allegheny West neighborhoods?
2. We see that there exists a clear correlation between the tree density and the overall benefits, however is this reflective of the reality? For example, given that Allegheny Center has the highest tree density where does it rank amongst the overall Air Quality Index score compared to the other neighborhoods?
3. Is there a correlation between the height of a tree species and its benefits? If yes, what is the type of correlation? Does the data imply whether neighborhood with tall trees are benefitted more or benefitted less compared to neighborhood with short trees? Can we derive insights from this data on whether planting tree species that are tall provide more benefits to the neighborhood or not, so that it will be useful for lanscaping and planning?
4. Why are there such big differences in the benefits provided across different neighborhoods for the same tree species? What factors could be contributing to this?

5. On the other hand, is this difference correlated to some of the neighborhoods that are at a disadvantage? If so, how much? In other words, is tree-inequality correlated to things like lower incomes or lower levels of educational attainment? Is tree equality across neighborhoods something that urban planners and policymakers should focus on?

6. Is there a correlation between the tree density and the wellbeing of individuals in a neighborhood?

7. Can we use different neighborhood features (including income, education, industrialization, population density, and many more) to predict the tree density in a given neighborhood?

Exploratory Data Analysis

Neighborhood Level Tree Density Data

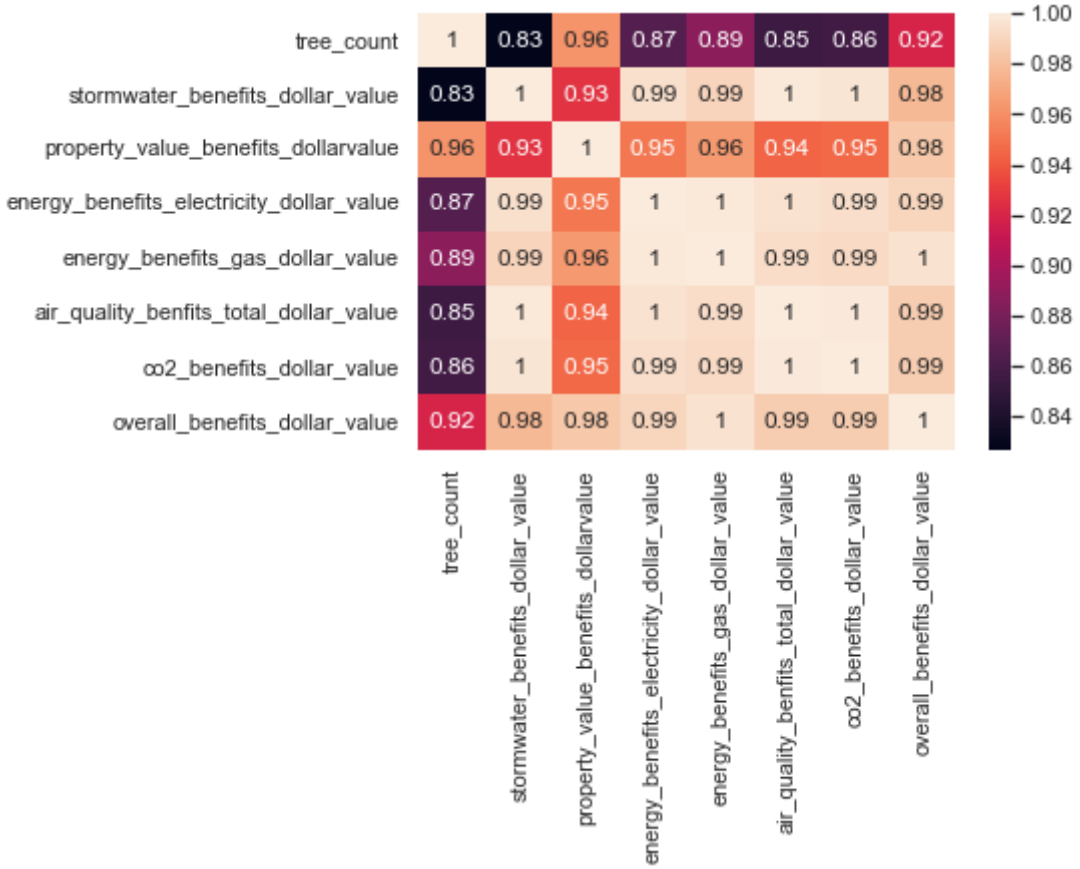
In [5]:
```
combined_data = pd.read_csv("cleaned_data/tree_density_data.csv")
tree_density_map = combined_data[['neighborhood', 'tree_count']].copy()
```

In [6]:
```
combined_data.columns
```

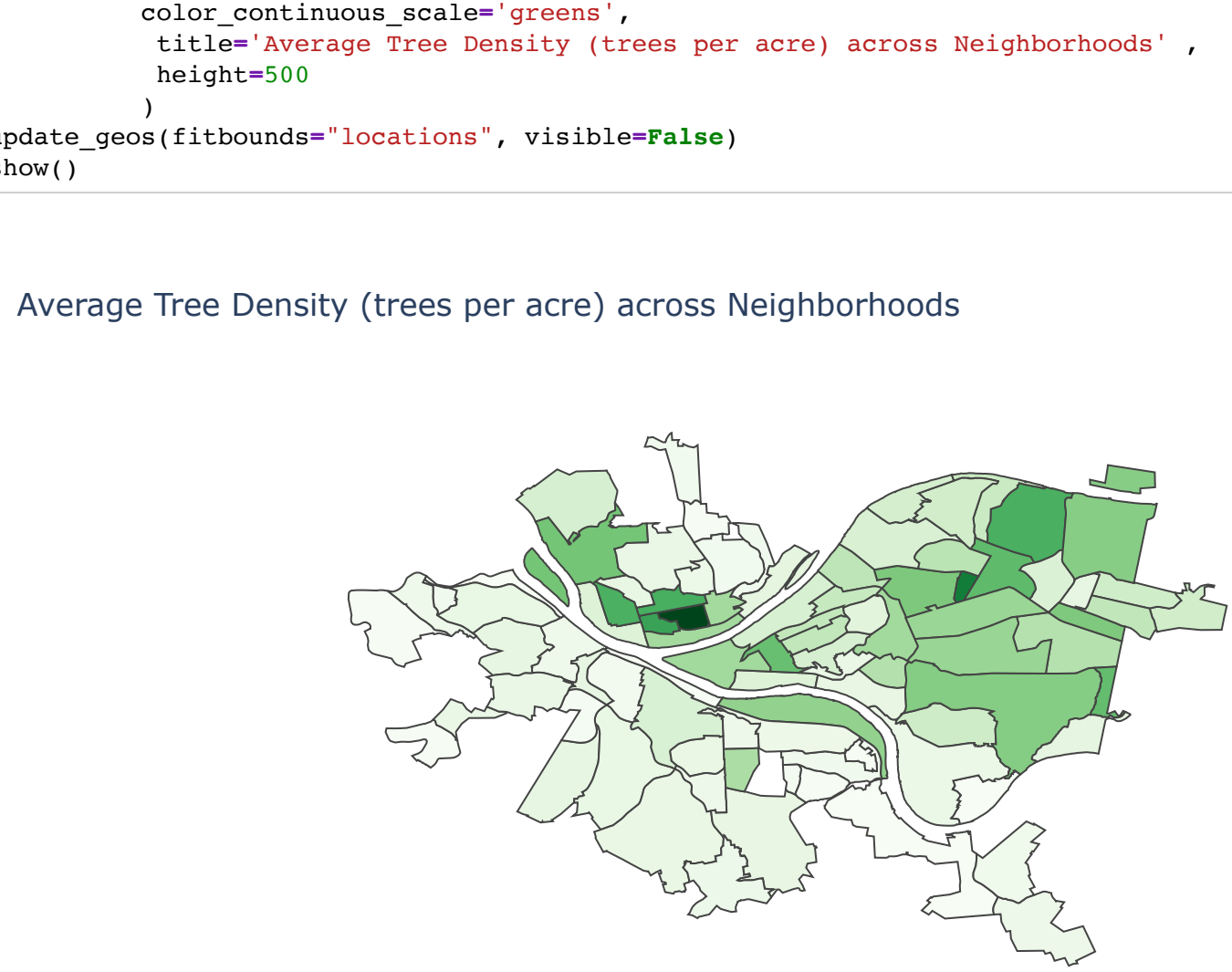
Out[6]:
```
Index(['Unnamed: 0', 'neighborhood', 'tree_count',
       'stormwater_benefits_dollar_value',
       'property_value_benefits_dollarvalue',
       'energy_benefits_electricity_dollar_value',
       'energy_benefits_gas_dollar_value',
       'air_quality_benfits_total_dollar_value', 'co2_benefits_dollar_value',
       'overall_benefits_dollar_value', 'Neighborhood_2010_AREA',
       'Neighborhood_2010_ACRES'],
      dtype='object')
```

In [7]:
```
# plot heatmap to observe correlationship between tree benefits
info = combined_data.drop(labels = ['Unnamed: 0', 'Neighborhood_2010_AREA', 'Neighborhood_2010_ACRES'],
corrMatrix = info.corr()
sns.heatmap(corrMatrix, annot=True)
```

Out[7]: <AxesSubplot:>



In [8]:
```
fig=px.choropleth(tree_density_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='tree_count',
                   color_continuous_scale='greens',
                    title='Average Tree Density (trees per acre) across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with Highest Tree Density

In [9]:
```
tree_density_map.sort_values('tree_count', ascending=False).head(5)
```

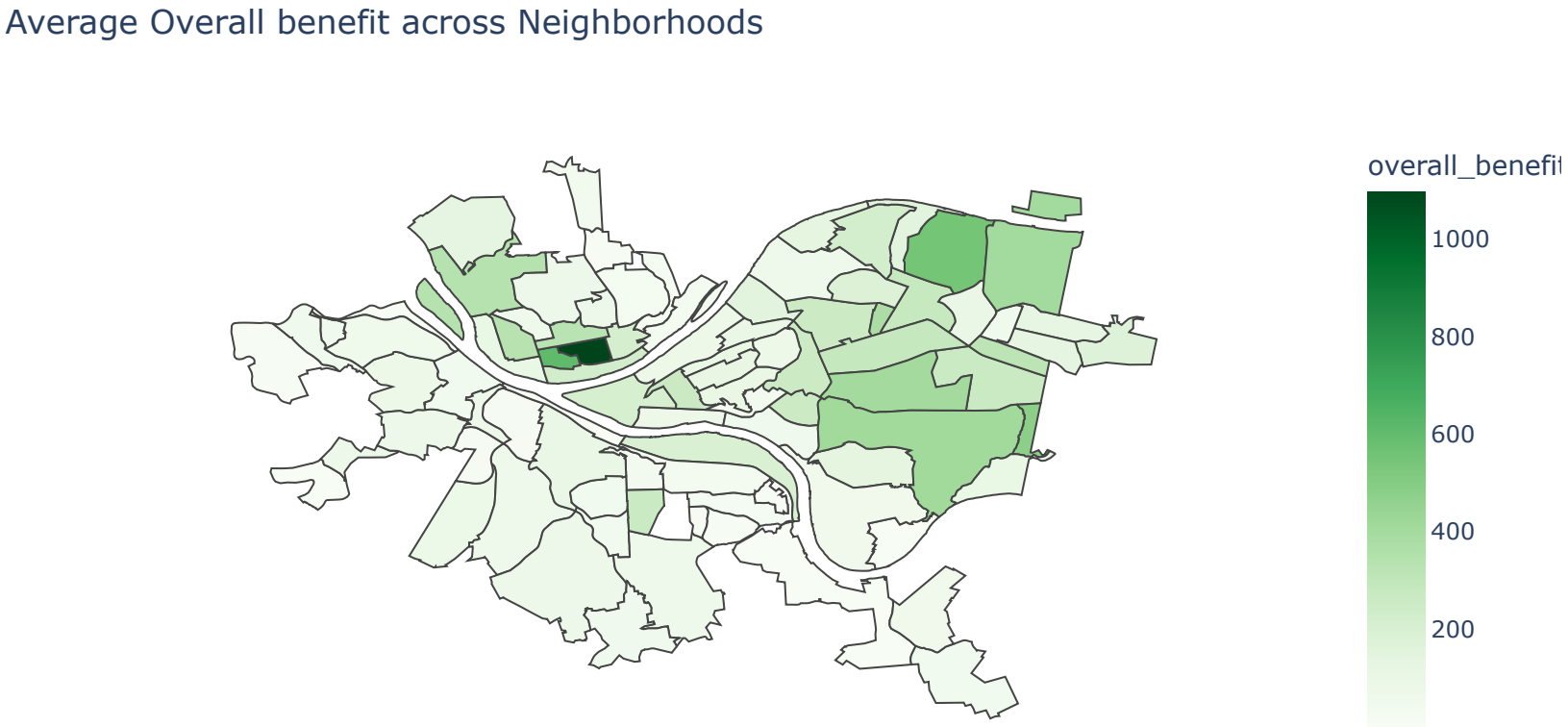
Out[9]:

	neighborhood	tree_count
0	Allegheny Center	6.406687
33	Friendship	5.183244
1	Allegheny West	4.227166
18	Central Northside	3.898483
39	Highland Park	3.861221

Neighborhood Level Overall Tree Benefits Data

In [10]:
```
overall_benefit_map = combined_data[['neighborhood', 'overall_benefits_dollar_value']].copy()
```

In [11]:
```
fig=px.choropleth(overall_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='overall_benefits_dollar_value',
                   color_continuous_scale='greens',
                    title='Average Overall benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest overall benefits

In [12]:
```
overall_benefit_map.sort_values('overall_benefits_dollar_value', ascending=False).head(5)
```

Out[12]:

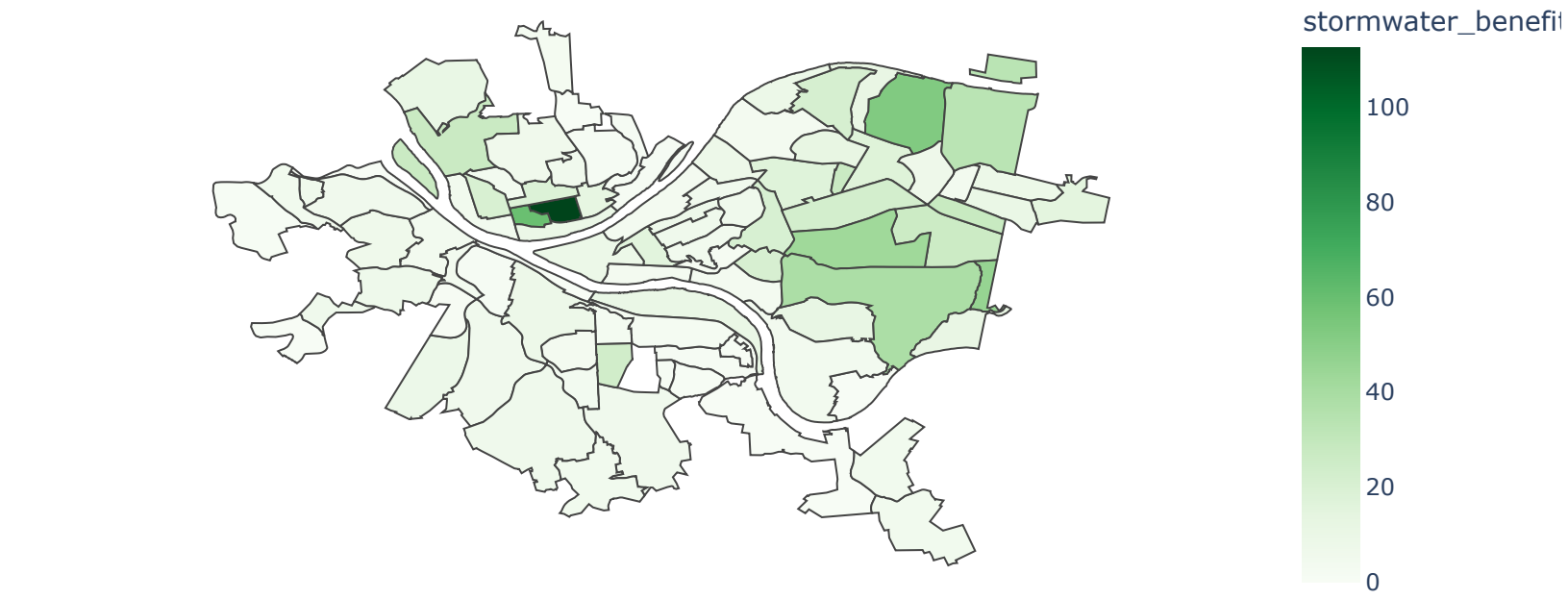
	neighborhood	overall_benefits_dollar_value
0	Allegheny Center	1098.066189
1	Allegheny West	603.947056
39	Highland Park	544.767046
65	Regent Square	477.596975
76	Squirrel Hill South	407.285619

Inference: 3 of the top 5 neighborhoods with the highest overall benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree Stormwater Benefits Data

In [13]:
```
stormwater_benefit_map = combined_data[['neighborhood', 'stormwater_benefits_dollar_value']].copy()
```

In [14]:
```
fig=px.choropleth(stormwater_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='stormwater_benefits_dollar_value',
                   color_continuous_scale='greens',
                    title='Average Stormwater benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest stormwater benefits

In [15]:
```
stormwater_benefit_map.sort_values('stormwater_benefits_dollar_value', ascending=False).head(5)
```

Out[15]:

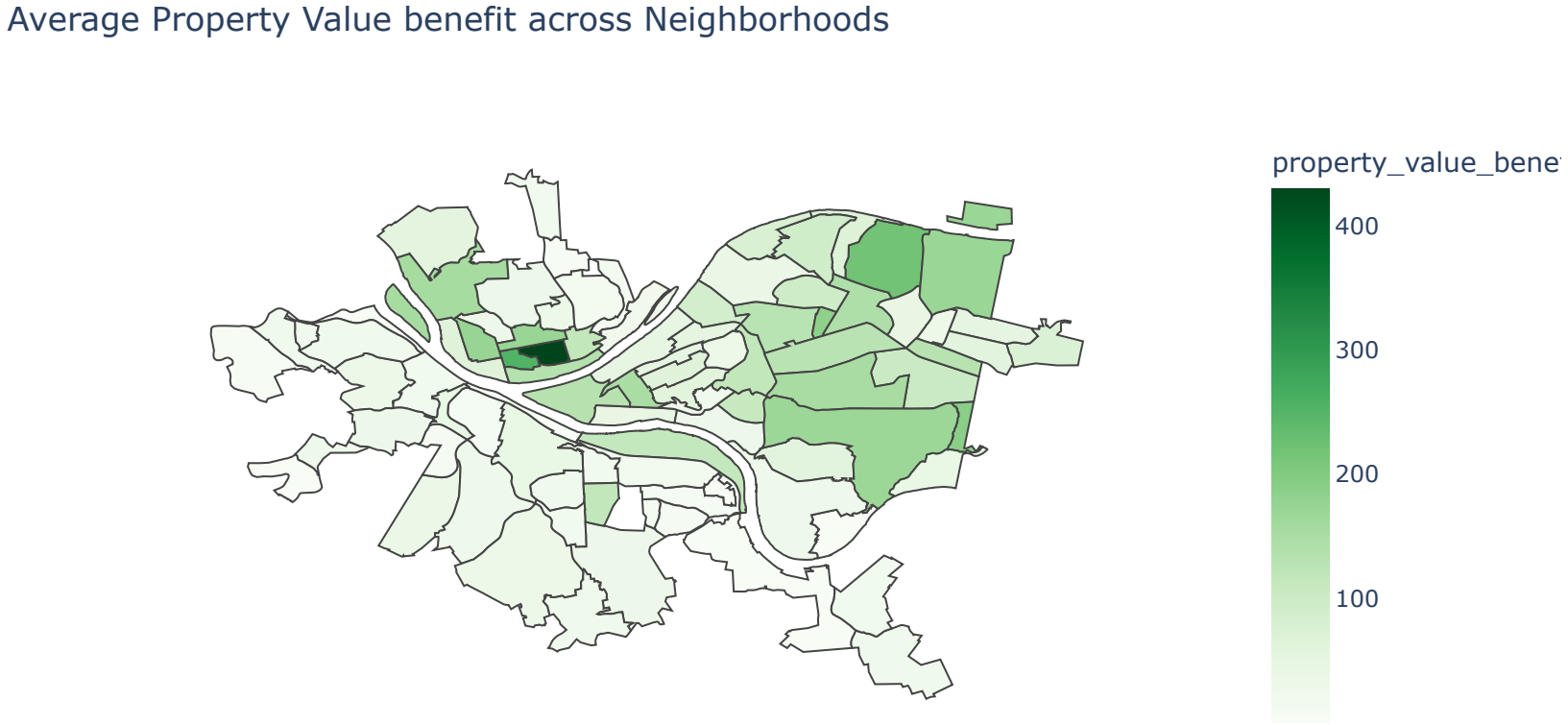
	neighborhood	stormwater_benefits_dollar_value
0	Allegheny Center	112.681320
1	Allegheny West	59.060754
39	Highland Park	52.148217
65	Regent Square	45.628699
75	Squirrel Hill North	42.471383

Inference: 3 of the top 5 neighborhoods with the highest stormwater benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree Property Value Benefits Data

In [16]:
```
property_value_benefit_map = combined_data[['neighborhood', 'property_value_benefits_dollarvalue']].copy
```

In [17]:
```
fig=px.choropleth(property_value_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='property_value_benefits_dollarvalue',
                   color_continuous_scale='greens',
                    title='Average Property Value benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest property value benefits

In [18]:
```
property_value_benefit_map.sort_values('property_value_benefits_dollarvalue', ascending=False).head(5)
```

Out[18]:

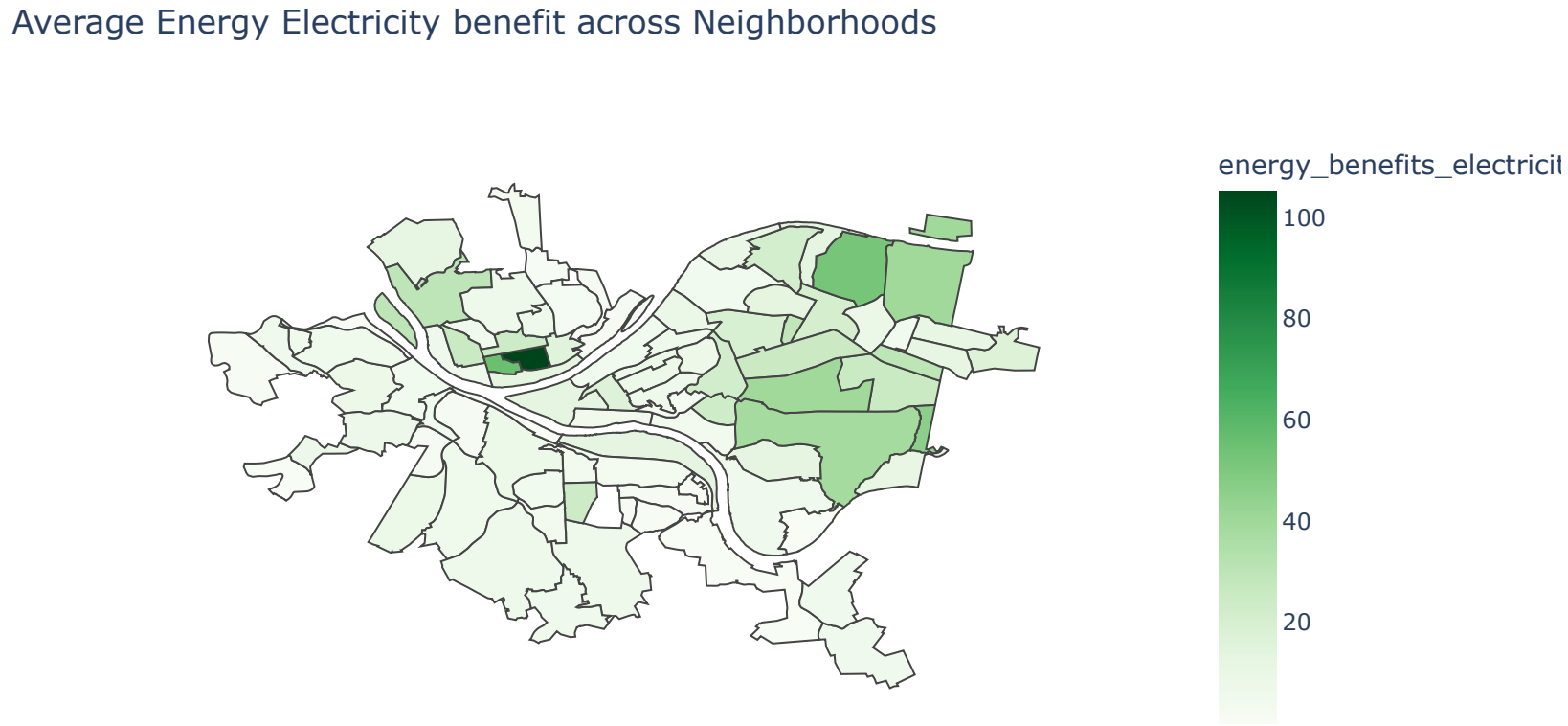
	neighborhood	property_value_benefits_dollarvalue
0	Allegheny Center	430.303156
1	Allegheny West	252.741943
39	Highland Park	217.631740
65	Regent Square	188.761117
33	Friendship	182.771638

Inference: 4 of the top 5 neighborhoods with the highest property value benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree Energy (Electricity) Benefits Data

In [19]:
```
energy_electricity_benefit_map = combined_data[['neighborhood', 'energy_benefits_electricity_dollar_valu
```

In [20]:
```
fig=px.choropleth(energy_electricity_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='energy_benefits_electricity_dollar_value',
                   color_continuous_scale='greens',
                    title='Average Energy Electricity benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest energy (electricity) benefits

In [21]:
```
energy_electricity_benefit_map.sort_values('energy_benefits_electricity_dollar_value', ascending=False).
```

Out[21]:

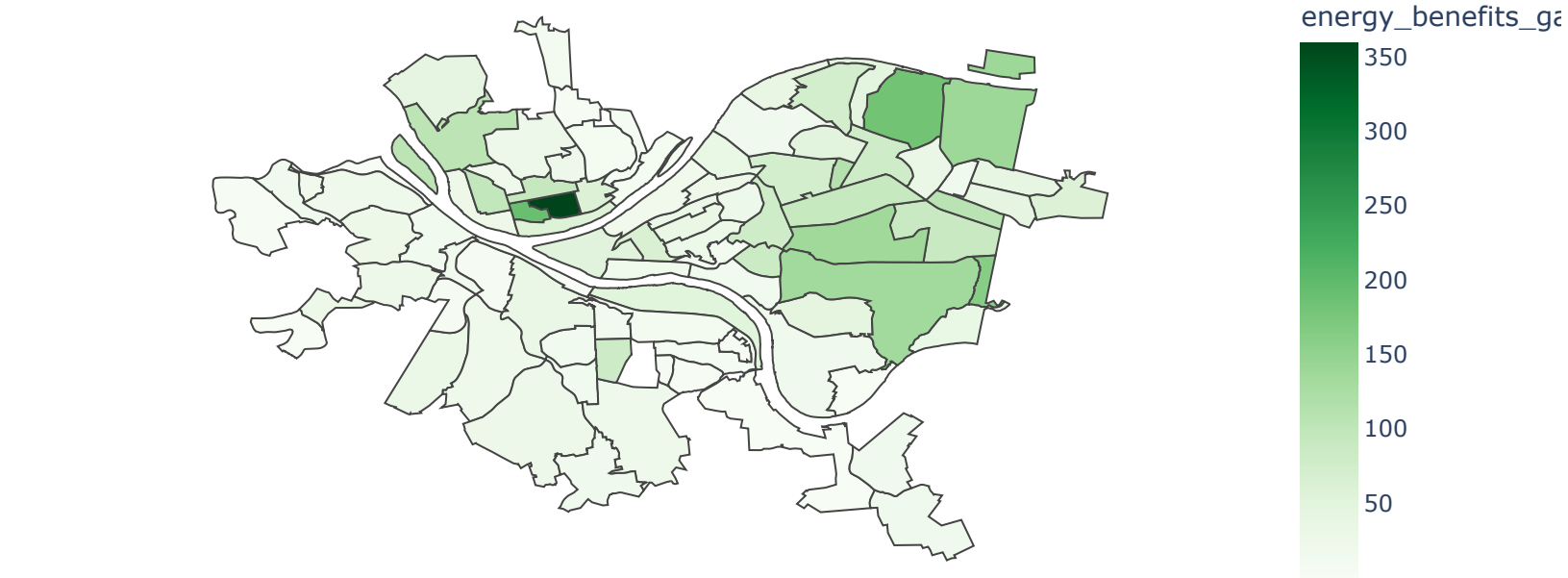
	neighborhood	energy_benefits_electricity_dollar_value
0	Allegheny Center	105.368577
1	Allegheny West	55.284105
39	Highland Park	52.100405
65	Regent Square	45.653537
75	Squirrel Hill North	39.547891

Inference: 3 of the top 5 neighborhoods with the highest energy (electricity) benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree Energy (Gas) Benefits Data

In [22]:
```
energy_gas_benefit_map = combined_data[['neighborhood', 'energy_benefits_gas_dollar_value']].copy()
```

In [23]:
```
fig=px.choropleth(energy_gas_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='energy_benefits_gas_dollar_value',
                   color_continuous_scale='greens',
                    title='Average Energy Gas benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest energy (gas) benefits

In [24]:
```
energy_gas_benefit_map.sort_values('energy_benefits_gas_dollar_value', ascending=False).head(5)
```

Out[24]:

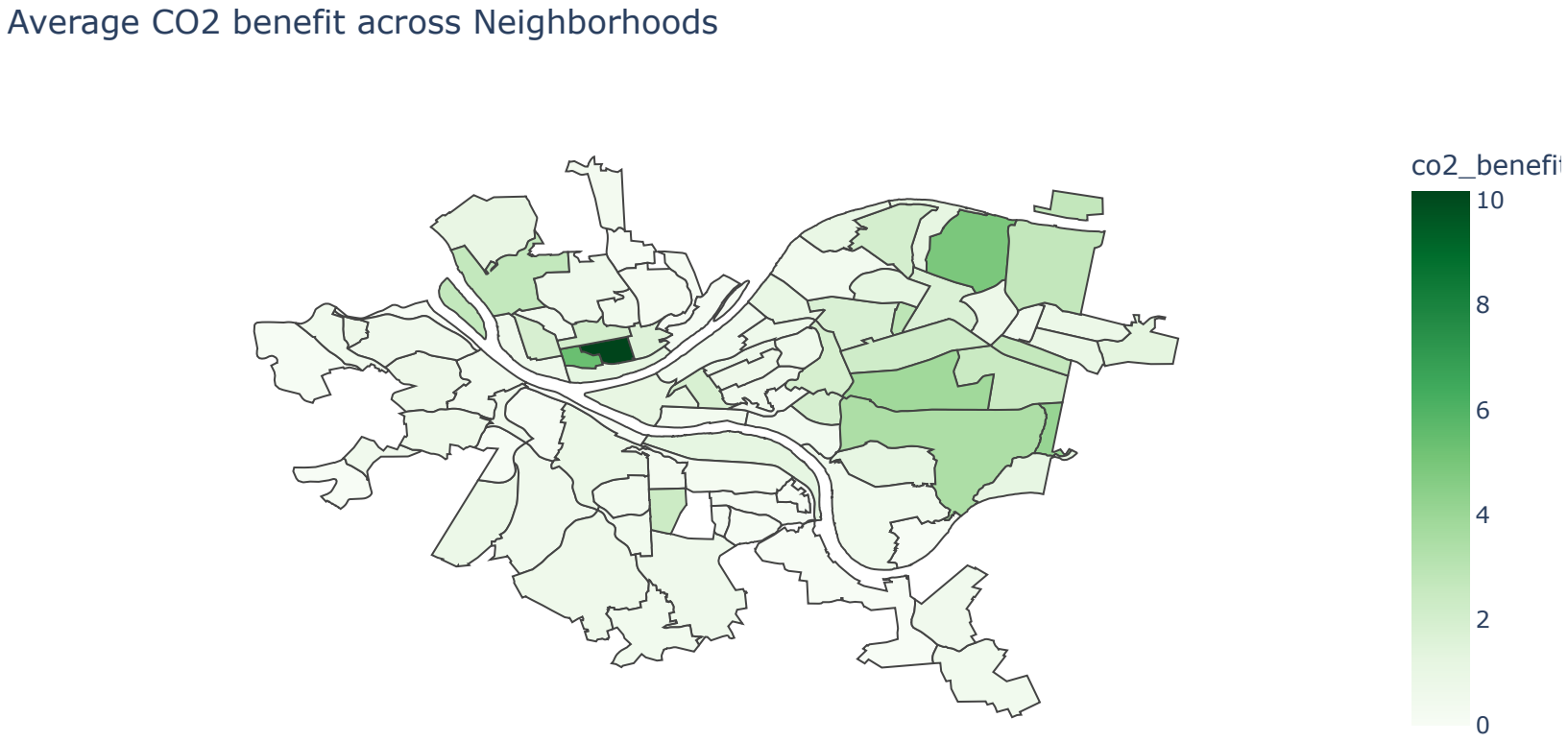
	neighborhood	energy_benefits_gas_dollar_value
0	Allegheny Center	359.936222
1	Allegheny West	190.093803
39	Highland Park	180.978794
65	Regent Square	161.498688
46	Lincoln-Lemington-Belmar	137.901237

Inference: 3 of the top 5 neighborhoods with the highest energy (gas) benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree CO2 Benefits Data

In [25]:
```
co2_benefit_map = combined_data[['neighborhood', 'co2_benefits_dollar_value']].copy()
```

In [26]:
```
fig=px.choropleth(co2_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='co2_benefits_dollar_value',
                   color_continuous_scale='greens',
                    title='Average CO2 benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest CO2 benefit

In [27]:
```
co2_benefit_map.sort_values('co2_benefits_dollar_value', ascending=False).head(5)
```

Out[27]:

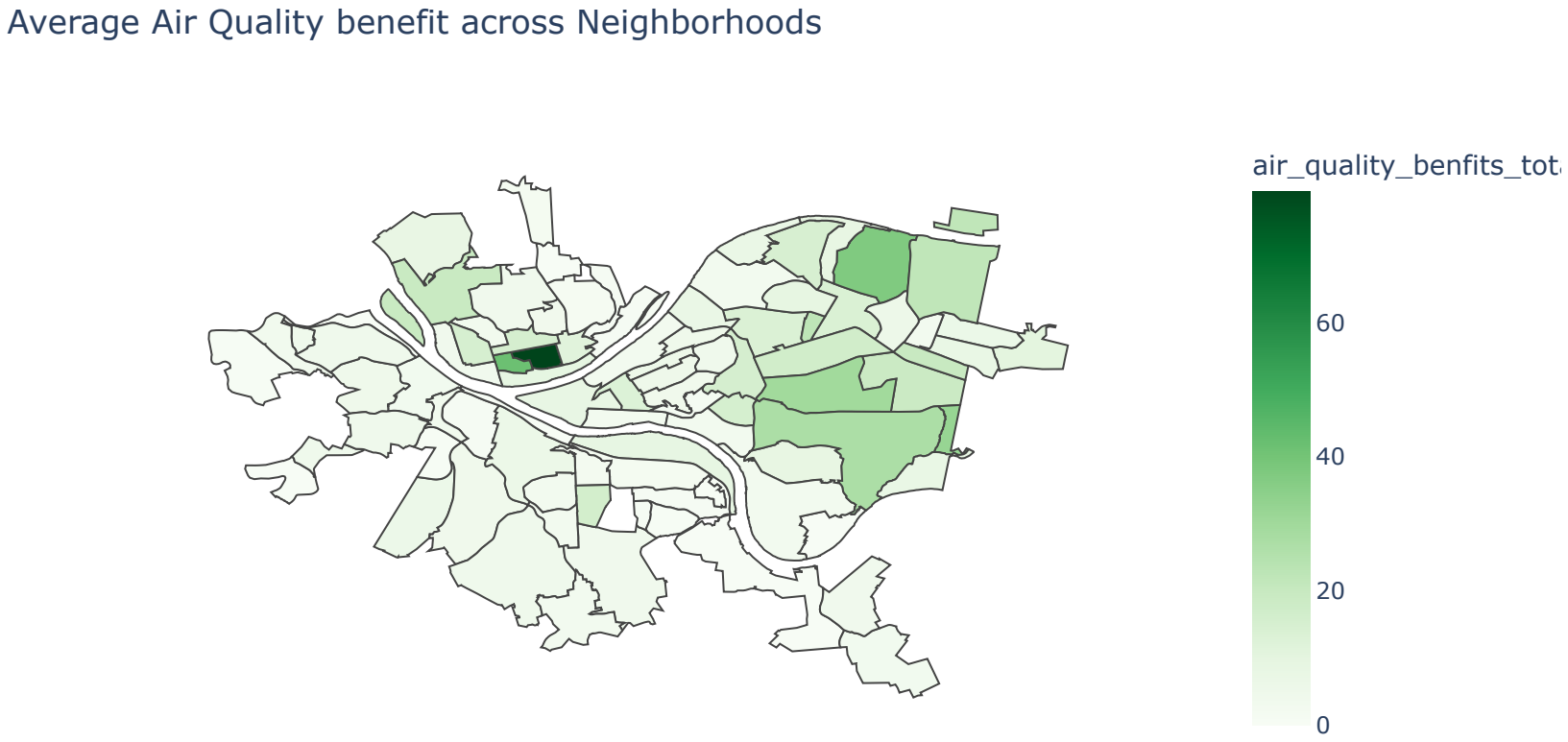
	neighborhood	co2_benefits_dollar_value
0	Allegheny Center	10.172134
1	Allegheny West	5.328917
39	Highland Park	4.885688
65	Regent Square	4.091595
75	Squirrel Hill North	3.799725

Inference: 3 of the top 5 neighborhoods with the highest CO2 benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Neighborhood Level Tree Air Quality Benefits Data

In [28]:
```
air_quality_benefit_map = combined_data[['neighborhood', 'air_quality_benfits_total_dollar_value']].copy
```

In [29]:
```
fig=px.choropleth(air_quality_benefit_map,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='air_quality_benfits_total_dollar_value',
                   color_continuous_scale='greens',
                    title='Average Air Quality benefit across Neighborhoods' ,
                    height=500
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```



Top 5 neighborhoods with highest Air quality benefit

In [30]:
```
air_quality_benefit_map.sort_values('air_quality_benfits_total_dollar_value', ascending=False).head(5)
```

Out[30]:

	neighborhood	air_quality_benfits_total_dollar_value
0	Allegheny Center	79.604781
1	Allegheny West	41.437534
39	Highland Park	37.170109
65	Regent Square	31.963339
75	Squirrel Hill North	29.528304

Inference: 3 of the top 5 neighborhoods with the highest air quality benefits are also in the top 5 neighborhoods with highest tree density. This exhibhits a clear positive correlation between the two

Tree Characteristics

In [31]:
```
df_trees = pd.read_csv("cleaned_data/cleaned_tree_data_5.csv", encoding="ISO-8859-1", low_memory=False)
```

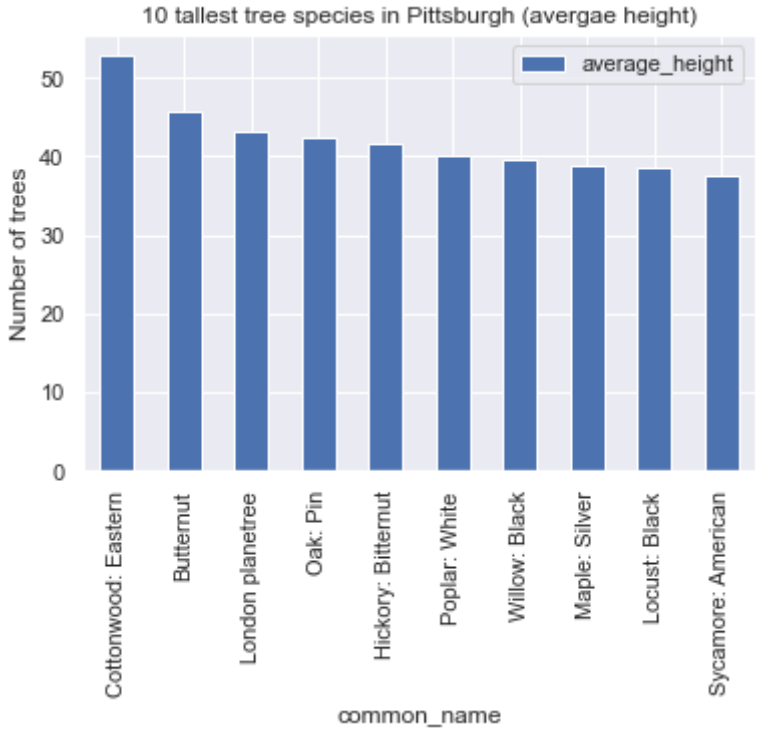
Average tree height by species

In [32]:
```
#Average height of tree species in increasing order
df_height_spec = df_trees.groupby("common_name")["height"].mean()
df_height_spec = df_height_spec.to_frame().reset_index()
df_height_spec.columns = ['common_name', 'average_height']
df_height_spec = df_height_spec.sort_values(by=['average_height'], ascending = False)
#Top 10 tallest trees species (average height) across pittsburgh
df_height_spec.head(10)
```

Out[32]:

	common_name	average_height
40	Cottonwood: Eastern	52.800000
25	Butternut	45.833333
103	London planetree	43.155996
142	Oak: Pin	42.351203
81	Hickory: Bitternut	41.632653
178	Poplar: White	40.000000
222	Willow: Black	39.600000
122	Maple: Silver	38.936686
102	Locust: Black	38.511367
208	Sycamore: American	37.615631

In [33]:
```
df_tallest_10 = df_height_spec.head(10)
ax = df_tallest_10.plot.bar(x='common_name', y='average_height', rot='vertical', ylabel = 'Number of tre
```



Average Tree Height across Neighborhoods

Top 10 neighborhoods with the highest average tree height

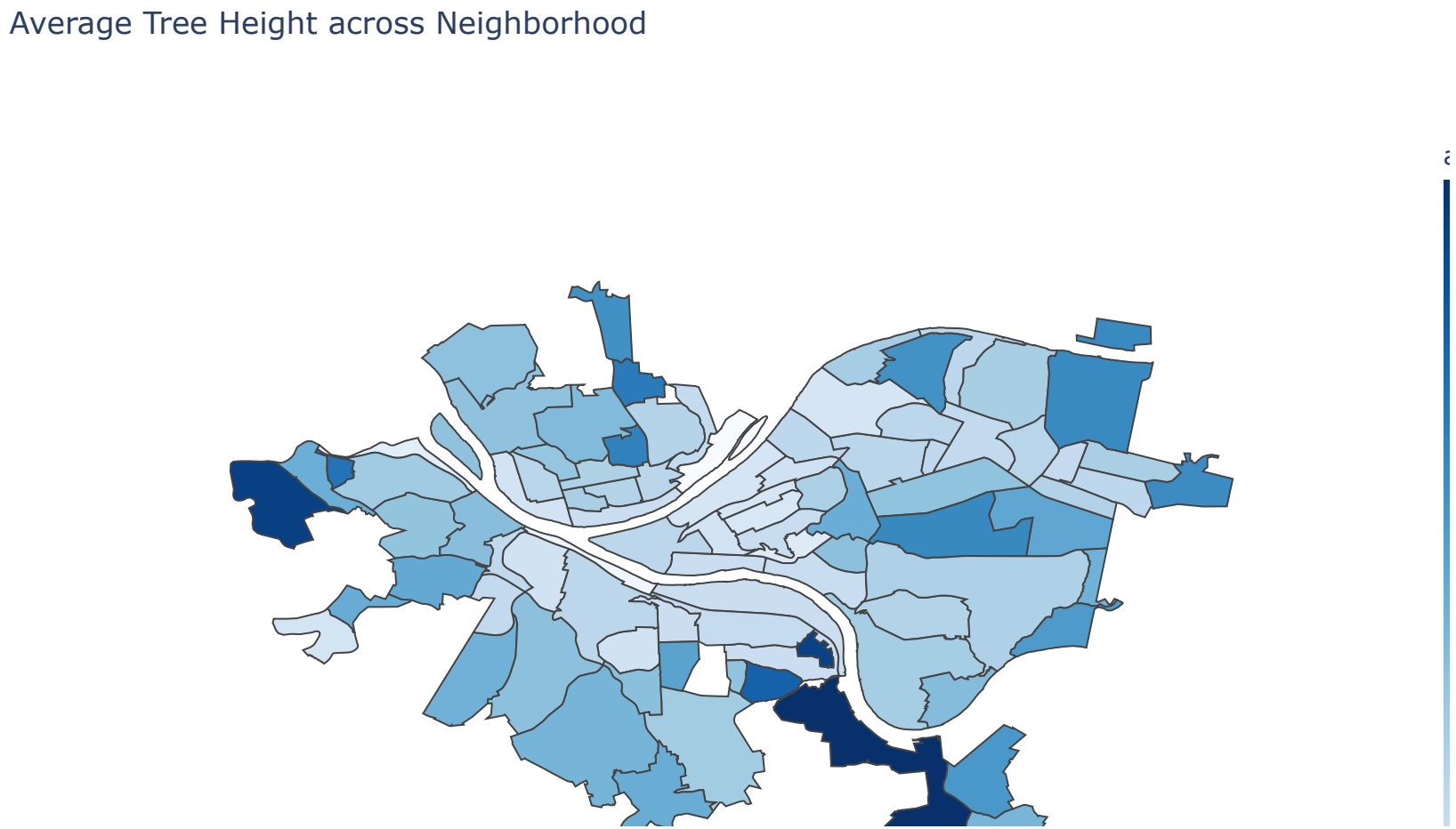
In [34]:
```
df_height = df_trees.groupby("neighborhood")["height"].mean()
df_height = df_height.to_frame().reset_index()
df_height.columns = ['neighborhood', 'average_height']
df_height1 = df_height.sort_values(by=['average_height'], ascending = False)
df_height1.head(10)
```

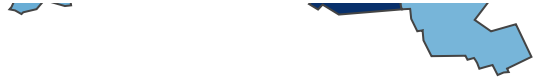
Out[34]:

	neighborhood	average_height
37	Hays	45.000000
31	Fairywood	42.964286
4	Arlington Heights	42.748561
77	St. Clair	38.500000
20	Chartiers City	36.318221
57	Northview Heights	35.333333
32	Fineview	34.524743
75	Squirrel Hill North	33.534065
46	Lincoln-Lemington-Belmar	33.336483
27	East Hills	33.173275

Choropleth map for average tree height across the neighborhoods of Pittsburgh

In [35]:
```
fig=px.choropleth(df_height,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='average_height',
                   color_continuous_scale='blues',
                    title='Average Tree Height across Neighborhood' ,
                    height=700
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
#img_bytes = fig.to_image(format="png")
#Image(img_bytes)
```





Average Tree Width across Neighborhoods

Top 10 neighborhoods with the highest average tree width

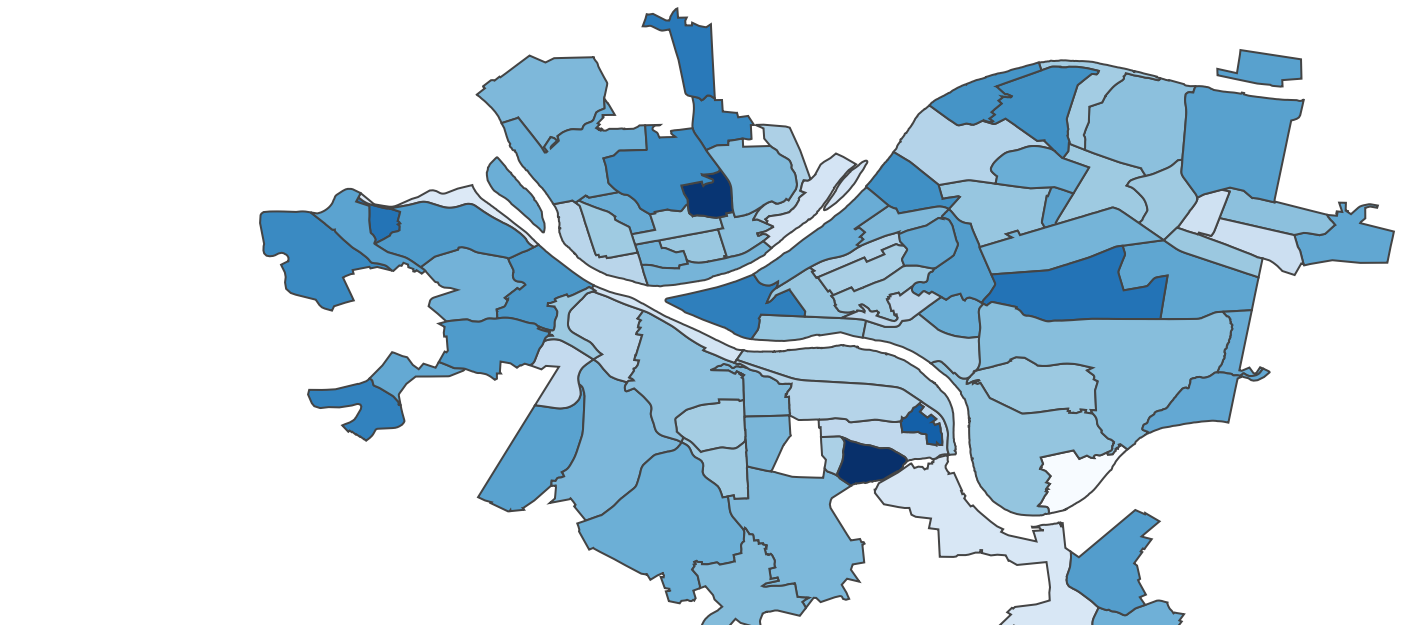
In [36]:
```
df_width = df_trees.groupby("neighborhood")["width"].mean()
df_width = df_width.to_frame().reset_index()
df_width.columns = ['neighborhood', 'average_width']
df_width1 = df_width.sort_values(by=['average_width'], ascending = False)
df_width1.head(10)
```

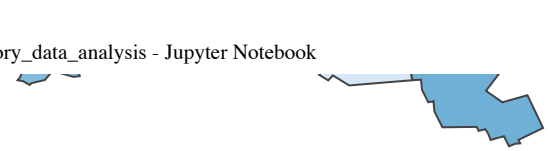
Out[36]:

	neighborhood	average_width
77	St. Clair	13.454069
32	Fineview	13.243513
4	Arlington Heights	11.448097
75	Squirrel Hill North	10.632508
20	Chartiers City	10.600609
80	Summer Hill	10.392978
16	Central Business District	10.152232
26	East Carnegie	10.027174
57	Northview Heights	9.750436
31	Fairywood	9.698687

Choropleth map for average tree width across the neighborhoods of Pittsburgh

In [37]:
```
fig=px.choropleth(df_width,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='average_width',
                   color_continuous_scale='blues',
                    title='Average Tree Width across Neighborhood' ,
                    height=700
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
#img_bytes = fig.to_image(format="png")
#Image(img_bytes)
```





Inference : 7 of the top 10 neighborhoods with the highest average tree height are also present in the top 10 neighborhoods with the highest average tree width. While there seems to be positive correlation between the average height and average width (tall trees tend to have more width), this can't be considered as a strong correlation. A strange observation is that the neighborhood 'Hays' has the highest average tree height in pittsburgh. However, it has one of the lowest average tree width in Pittsburgh!

The most prevalent species in each neighborhood

In [38]:
```
groupBySpeciesAndNeighborhood = df_trees.groupby(['neighborhood','common_name'])['id'].count()
groupBySpeciesAndNeighborhood = groupBySpeciesAndNeighborhood.to_frame().reset_index()
prevalent_species = groupBySpeciesAndNeighborhood.loc[groupBySpeciesAndNeighborhood.groupby(['neighborhc
prevalent_species
```

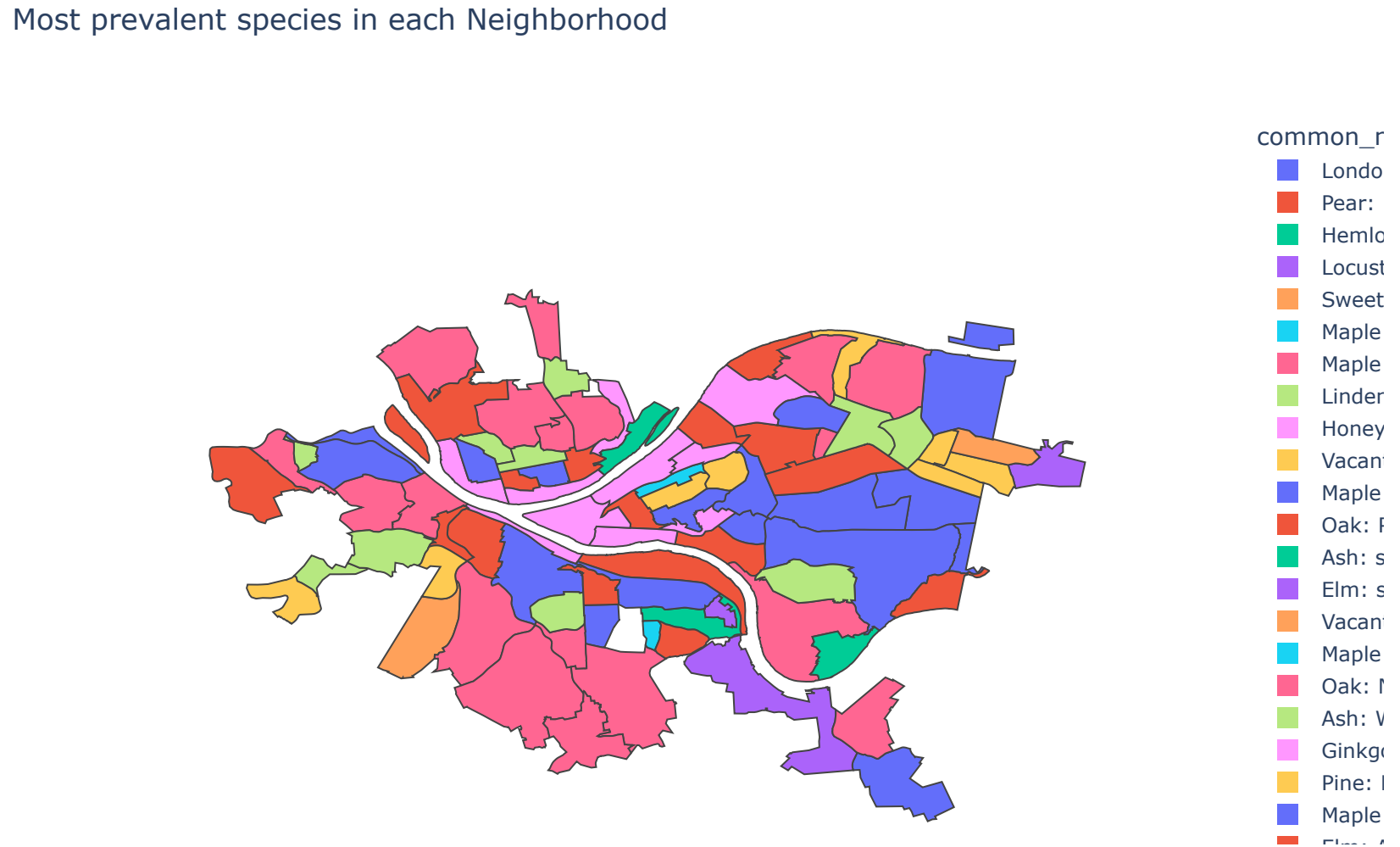
Out[38]:

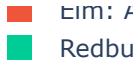
	neighborhood	common_name	id
0	Allegheny Center	London planetree	109
1	Allegheny West	Pear: Callery	44
2	Allentown	Pear: Callery	14
3	Arlington	Hemlock: Eastern	14
4	Arlington Heights	Locust: Black	2
...	...	...	...
85	Upper Lawrenceville	Pear: Callery	65
86	West End	Pear: Callery	46
87	West Oakland	Honeylocust: Thornless	17
88	Westwood	Linden: Littleleaf	29
89	Windgap	Maple: Norway	18

90 rows × 3 columns

Choropleth map showing the tree species that is the highest in number in each neighborhood

In [39]:
```
fig=px.choropleth(prevalent_species,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geo
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='common_name',
                   color_continuous_scale='Inferno',
                    title='Most prevalent species in each Neighborhood' ,
                    height=700
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.show()
```





The top 5 tree species that are the most prevalant in many neighborhoods

In [40]:
```
overall_prevalent = prevalent_species.groupby(['common_name'])['id'].count()
overall_prevalent = overall_prevalent.to_frame().reset_index()
overall_prevalent = overall_prevalent.sort_values(by=['id'], ascending = False)
overall_prevalent.head(5)
```

Out[40]:

	common_name	id
11	Maple: Norway	18
17	Pear: Callery	12
9	London planetree	10
7	Linden: Littleleaf	9
6	Honeylocust: Thornless	7

Inference: It can be seen that Maple:Norway is the most prevalent species in 18 neighborhoods, followed by Pear:Callery which is the most prevalent in 12 neighborhoods and then London planetree which is the most prevalent in 10 neighborhoods and so on.

Distribution of species across neighborhoods

In [41]:
```
#Selecting tree species whose count is more than 50
df_trees_thresh = df_trees.groupby(['common_name'])['id'].count()
df_trees_thresh1 = df_trees_thresh.to_frame().reset_index()
df_trees_thresh1 = df_trees_thresh1.sort_values(by=['id'], ascending = False)
df_trees_thresh1 = df_trees_thresh1[df_trees_thresh1['id'] > 50]
df_trees_thresh1
```

Out[41]:

	common_name	id
118	Maple: Norway	3717
120	Maple: Red	3421
103	London planetree	3224
156	Pear: Callery	2969
217	Vacant Site Small	2418
...	...	...
190	Serviceberry: spp.	66
202	Spruce: White	59
87	Hophornbeam: American	58
143	Oak: Sawtooth	57
106	Magnolia: Saucer	55

79 rows × 2 columns

Run the following cell and select a species from the dropdown menu and run the following cells to visualize the distribution of the selected species across the neighborhoods of Pittsburgh.

In [42]:
```
#Run this cell
#Select a species from the dropdown menu
#Run the following cells
selected_spec = widgets.Dropdown(options = df_trees_thresh1.common_name, value='Maple: Norway', descript
print("\nSelect a species\n")
display(selected_spec)
```

```
Select a species
```

Species: Maple: Norway

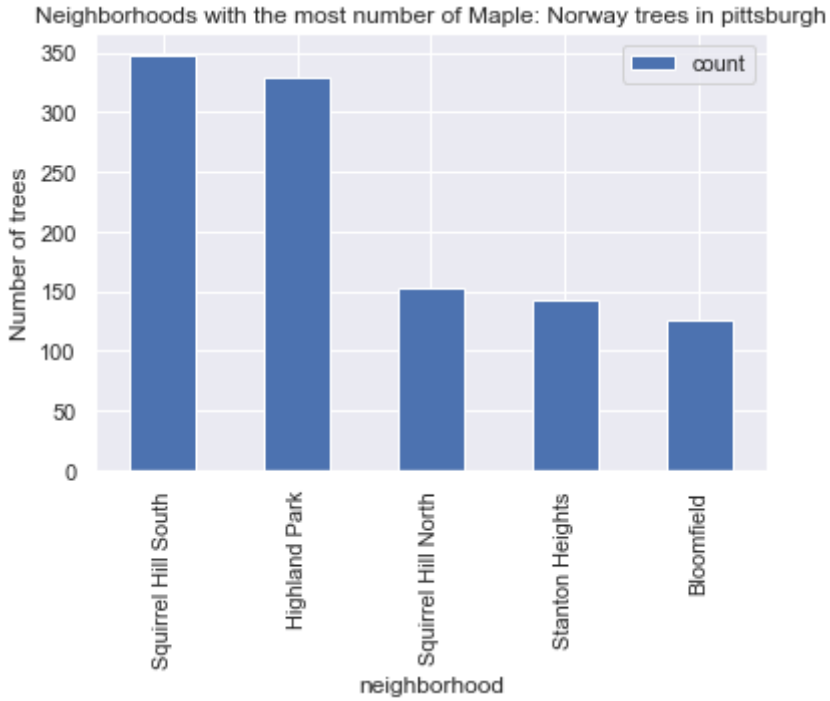
In [43]:
```
df_spec = df_trees.groupby(['common_name', 'neighborhood'])['id'].count()
df_spec = df_spec.to_frame().reset_index()
df_spec = df_spec.loc[df_spec['common_name'] == selected_spec.value]
#df_spec['percent'] = ((df_spec['id']/df_spec['id'].sum()) * 100)
df_spec = df_spec.rename(columns={"id": "count"})
df_spec
```

Out[43]:

	common_name	neighborhood	count
1968	Maple: Norway	Allegheny Center	48
1969	Maple: Norway	Allegheny West	21
1970	Maple: Norway	Allentown	9
1971	Maple: Norway	Arlington	9
1972	Maple: Norway	Banksville	39
...	...	...	...
2041	Maple: Norway	Upper Hill	18
2042	Maple: Norway	Upper Lawrenceville	26
2043	Maple: Norway	West End	12
2044	Maple: Norway	Westwood	26
2045	Maple: Norway	Windgap	18

78 rows × 3 columns

In [44]:
```
str1 = "Neighborhoods with the most number of " + str(selected_spec.value) + " trees in pittsburgh"
df_spec1 = df_spec.sort_values(by=['count'], ascending = False)
df_spec1 = df_spec1.head(5)
ax = df_spec1.plot.bar(x='neighborhood', y='count', rot='vertical', ylabel = 'Number of trees', title= s
```



In [45]:
```
full_neigh = pd.DataFrame({'neighborhood' : df_trees['neighborhood'].unique()})
full_neigh = full_neigh.rename(columns={"id": "count"})
full_neigh = full_neigh.merge(df_spec, how = 'outer', on = ['neighborhood'])
full_neigh['count'] = full_neigh['count'].fillna(0)
full_neigh = full_neigh.drop(labels = ['common_name'], axis = 1)
full_neigh.sort_values(['count'], ascending= False)
```

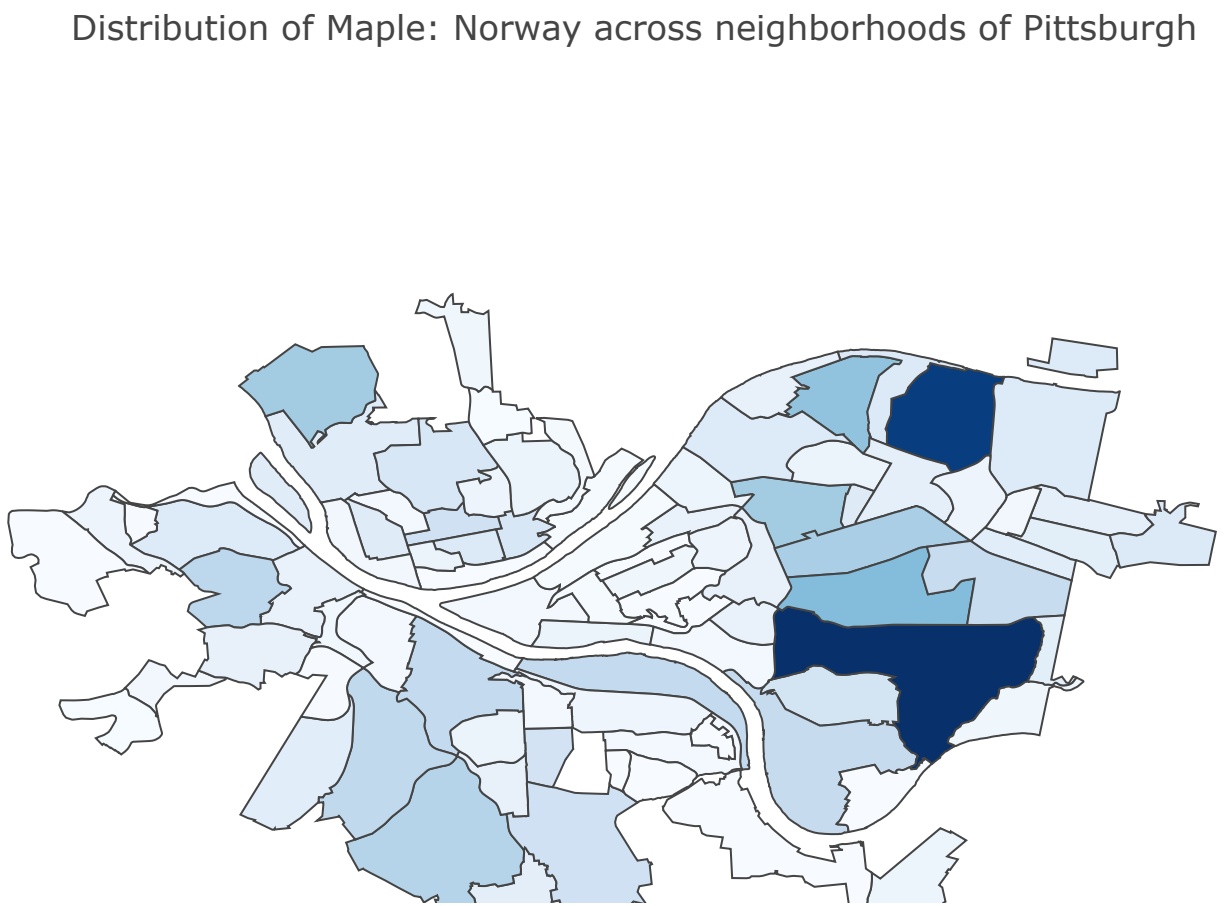
Out[45]:

	neighborhood	count
26	Squirrel Hill South	348.0
4	Highland Park	329.0
42	Squirrel Hill North	153.0
5	Stanton Heights	142.0
28	Brighton Heights	125.0
...	...	...
37	Bedford Dwellings	0.0
77	Fairywood	0.0
62	Terrace Village	0.0
35	Arlington Heights	0.0
89	Hays	0.0

90 rows × 2 columns

Choropleth map showing the distribution of the selected species across the neighborhoods of Pittsburgh

In [46]:
```
str2 = "Distribution of " + str(selected_spec.value) + " across neighborhoods of Pittsburgh"
fig=px.choropleth(full_neigh,
                  geojson="https://raw.githubusercontent.com/blackmad/neighborhoods/master/gn-pittsburgh.geoj
                  featureidkey='properties.name',
                  locations='neighborhood',        #column in dataframe
                  color='count',
                   color_continuous_scale= 'blues',
                    title= str2 ,
                    height=700,
                   )
fig.update_geos(fitbounds="locations", visible=False)
fig.layout.template = None
fig.show()
```



Tree Species Specific Benefits

The dataset being explored is the cleaned "City of Pittsburgh Trees" dataset. We will investigate the benefits (in terms of dollar value) of different tree species and try to identify some patterns. Then, some intra-species investigation will be done, as well as some neighborhood-level explorations.

In [47]:
```
df_trees = pd.read_csv("cleaned_data/cleaned_tree_data_5.csv", encoding="ISO-8859-1", low_memory=False)
tree_stat = df_trees.groupby(["common_name"]).agg(["count", "mean"]).reset_index()
```

Only selected trees species with more than 10 datapoints to reduce the effect of outliers.

In [48]:
```
tree_stat = tree_stat[tree_stat["id"]["count"] >= 10]
tree_stat.head(2)
```

Out[48]:

	common_name	id		height		width		growth_space_length		growth_space_width	...	public_works_di	
		count	mean	count	mean	count	mean	count	mean	count	...	count	mean
0	Amur Corktree	45	1.010417e+09	45	17.881481	45	5.013827	45	51.330370	45	...	45	2.2
1	Amur Maackia	17	1.158984e+09	17	13.550173	17	7.231834	17	10.401384	17	...	17	3.4

2 rows × 97 columns

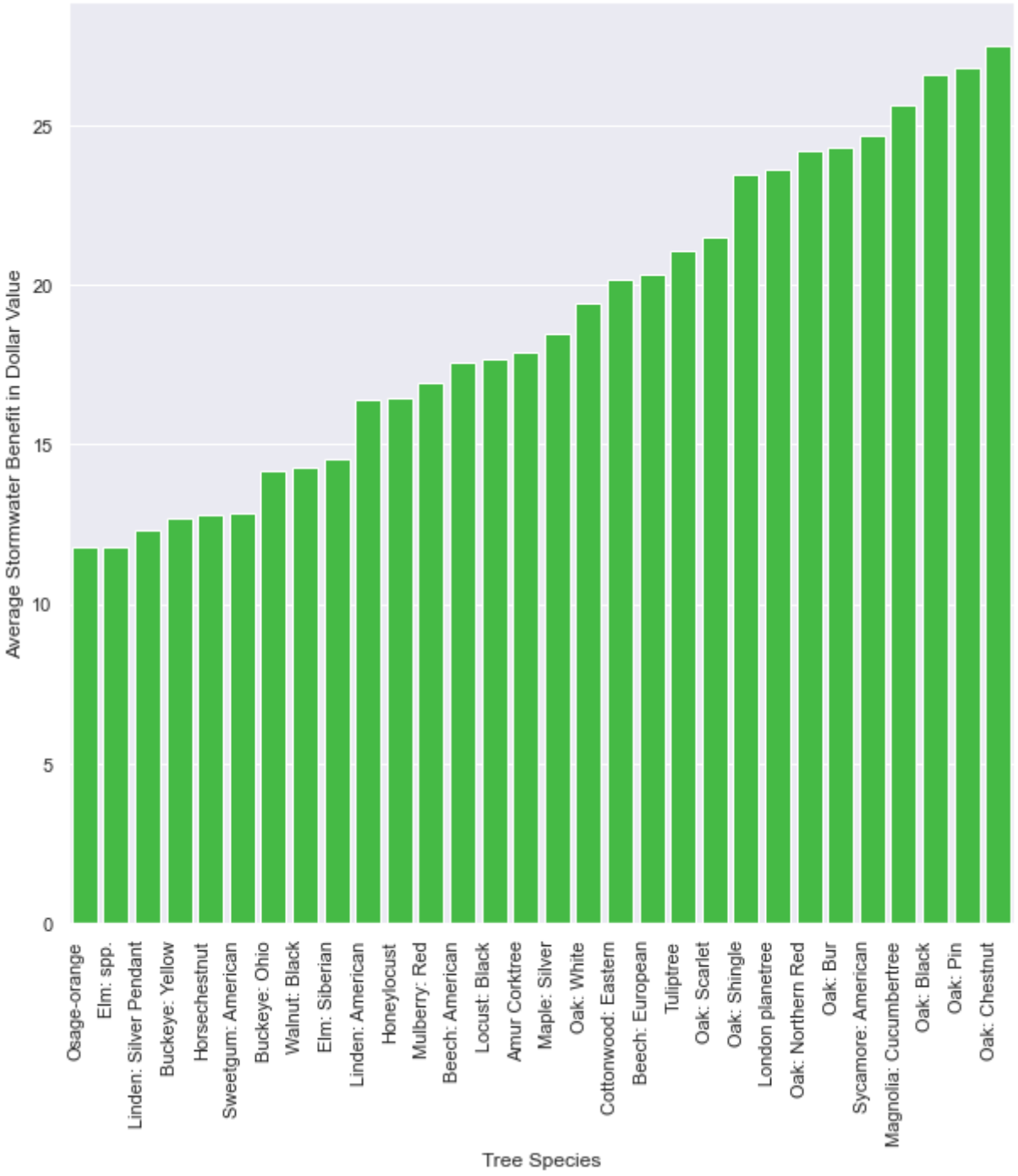
Since there are many different tree species, only top 30 will be shown for each benefit category. The common names, instead of the

scientific names, of the trees will be used since they are more recognizable.

Top 30 tree species that provide the most stormwater benefits

In [49]:
```
storm_water_benefit = tree_stat.sort_values([("stormwater_benefits_dollar_value","mean")])
storm_water_benefit = storm_water_benefit.tail(30)
```

In [50]:
```
storm_water_bar_plot = sns.barplot(x=storm_water_benefit["common_name"], y=storm_water_benefit["stormwat
storm_water_bar_plot.set_xticklabels(storm_water_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
storm_water_bar_plot.set_xlabel("Tree Species", fontsize = 12)
storm_water_bar_plot.set_ylabel("Average Stormwater Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```

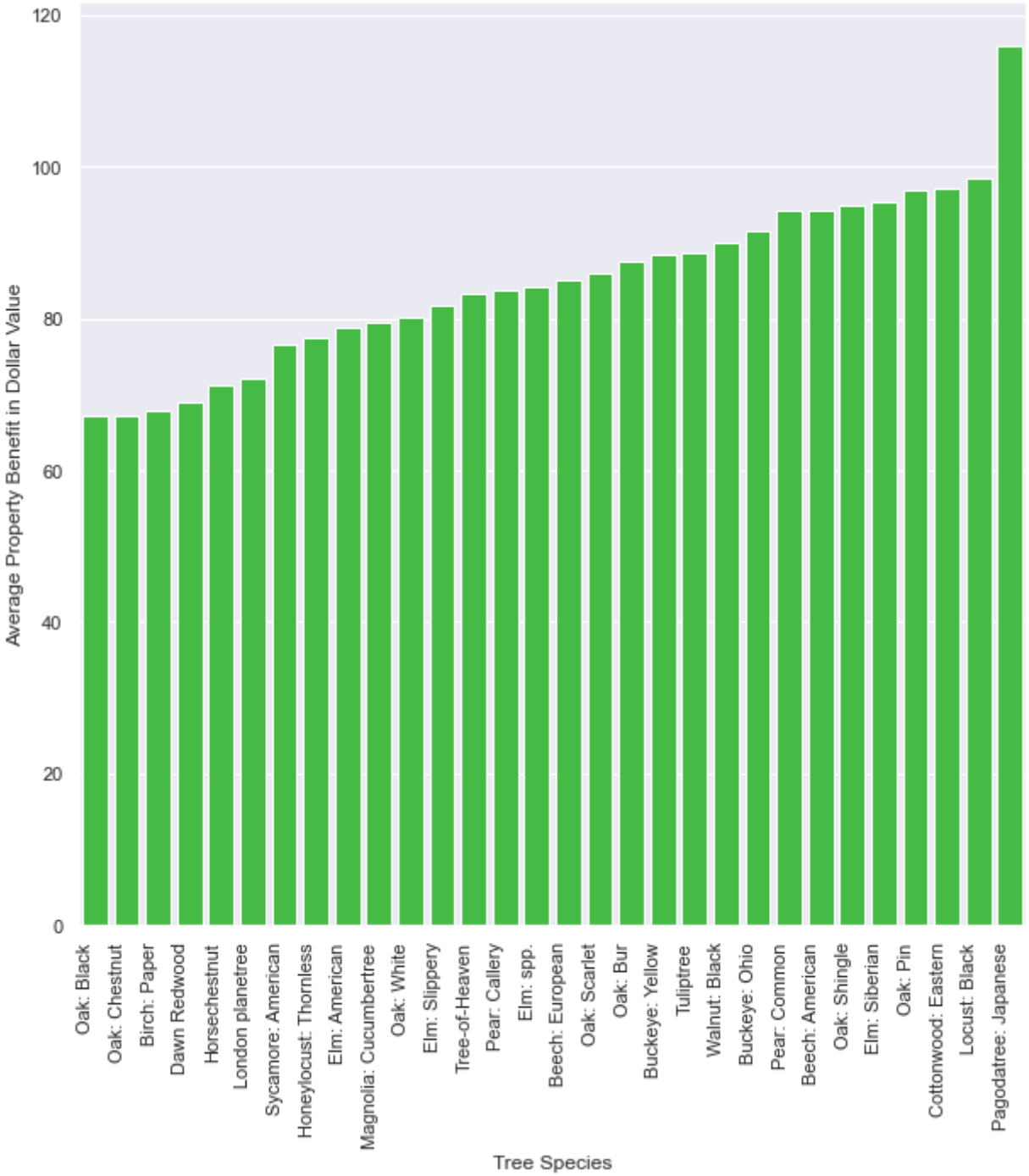


Different oak species offered the most benefits.

Top 30 tree species that provide the most property value benefits

In [51]:
```
property_benefit = tree_stat.sort_values([("property_value_benefits_dollarvalue","mean")])
property_benefit = property_benefit.tail(30)
```

In [52]:
```
property_benefit_bar_plot = sns.barplot(x=property_benefit["common_name"], y=property_benefit["property_
property_benefit_bar_plot.set_xticklabels(property_benefit_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
property_benefit_bar_plot.set_xlabel("Tree Species", fontsize = 12)
property_benefit_bar_plot.set_ylabel("Average Property Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```

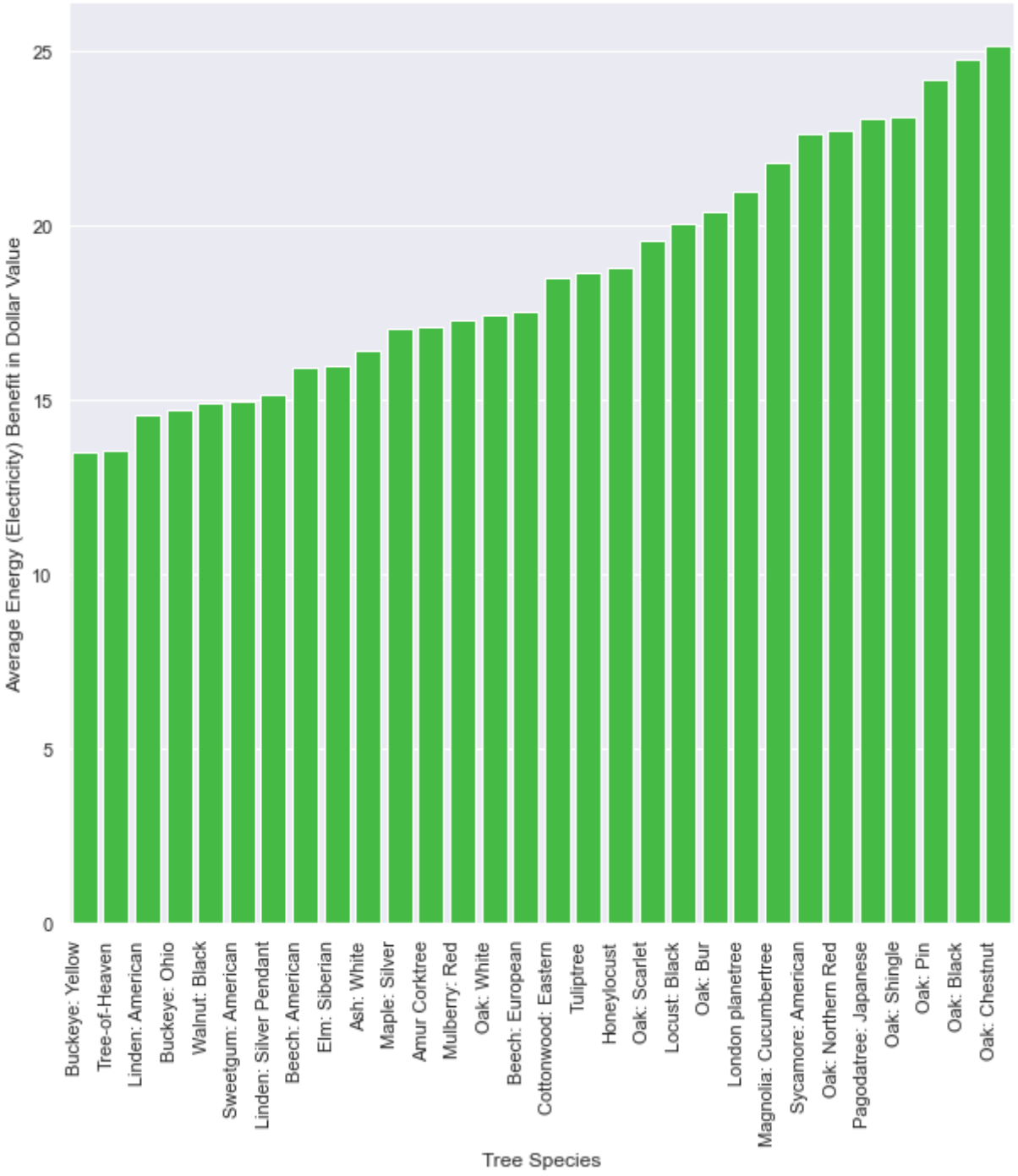

Average Property Benefit in Dollar Value
0
20
40
60
80
100
120
Oak: Black
Oak: Chestnut
Birch: Paper
Dawn Redwood
Horsechestnut
London planetree
Sycamore: American
Honeylocust: Thornless
Elm: American
Magnolia: Cucumbertree
Oak: White
Elm: Slippery
Tree-of-Heaven
Pear: Callery
Elm: spp.
Beech: European
Oak: Scarlet
Oak: Bur
Buckeye: Yellow
Tuliptree
Walnut: Black
Buckeye: Ohio
Pear: Common
Beech: American
Oak: Shingle
Elm: Siberian
Oak: Pin
Cottonwood: Eastern
Locust: Black
Pagodatree: Japanese
Tree Species

Japanese Pagodatree and black locust trees provide the most property value benefits.

Top 30 tree species that provide the most electricity saving benefits

In [53]:
```
energy_benefit_elec = tree_stat.sort_values([("energy_benefits_electricity_dollar_value","mean")])
energy_benefit_elec = energy_benefit_elec.tail(30)
```

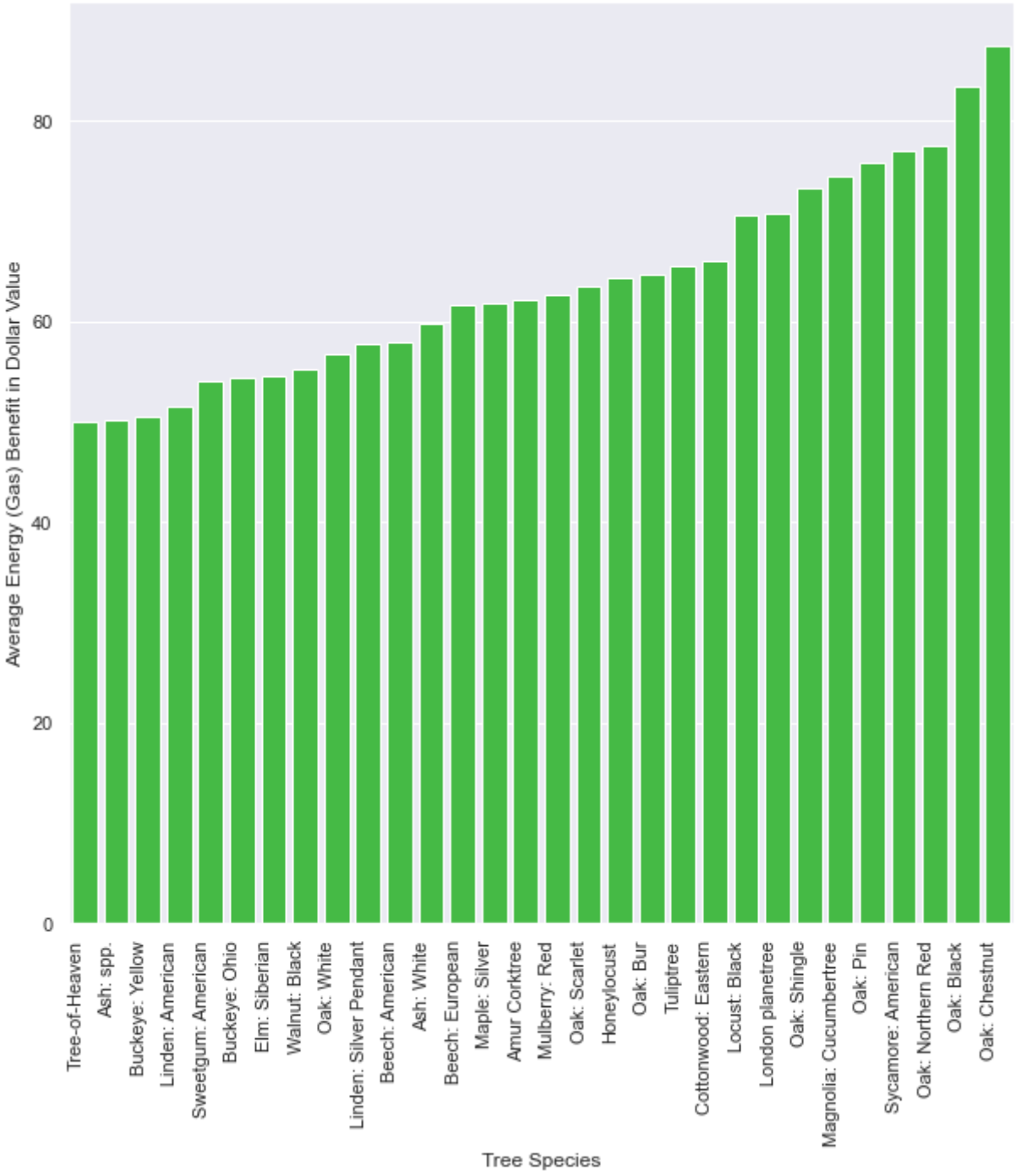
In [54]:
```
energy_benefit_elec_bar_plot = sns.barplot(x=energy_benefit_elec["common_name"], y=energy_benefit_elec['
energy_benefit_elec_bar_plot.set_xticklabels(energy_benefit_elec_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
energy_benefit_elec_bar_plot.set_xlabel("Tree Species", fontsize = 12)
energy_benefit_elec_bar_plot.set_ylabel("Average Energy (Electricity) Benefit in Dollar Value", fontsize
plt.gcf().set_size_inches(10,10)
```


Average Energy (Electricity) Benefit in Dollar Value
0
5
10
15
20
25
Buckeye: Yellow
Tree-of-Heaven
Linden: American
Buckeye: Ohio
Walnut: Black
Sweetgum: American
Linden: Silver Pendant
Beech: American
Elm: Siberian
Ash: White
Maple: Silver
Amur Corktree
Mulberry: Red
Oak: White
Beech: European
Cottonwood: Eastern
Tuliptree
Honeylocust
Oak: Scarlet
Locust: Black
Oak: Bur
London planetree
Magnolia: Cucumbertree
Sycamore: American
Oak: Northern Red
Pagodatree: Japanese
Oak: Shingle
Oak: Pin
Oak: Black
Oak: Chestnut
Tree Species

Top 30 tree species that provide the most gas saving benefits

In [55]:
```
energy_benefit_gas = tree_stat.sort_values([("energy_benefits_gas_dollar_value","mean")])
energy_benefit_gas = energy_benefit_gas.tail(30)
```

In [56]:
```
energy_benefit_gas_bar_plot = sns.barplot(x=energy_benefit_gas["common_name"], y=energy_benefit_gas["ene
energy_benefit_gas_bar_plot.set_xticklabels(energy_benefit_gas_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
energy_benefit_gas_bar_plot.set_xlabel("Tree Species", fontsize = 12)
energy_benefit_gas_bar_plot.set_ylabel("Average Energy (Gas) Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```

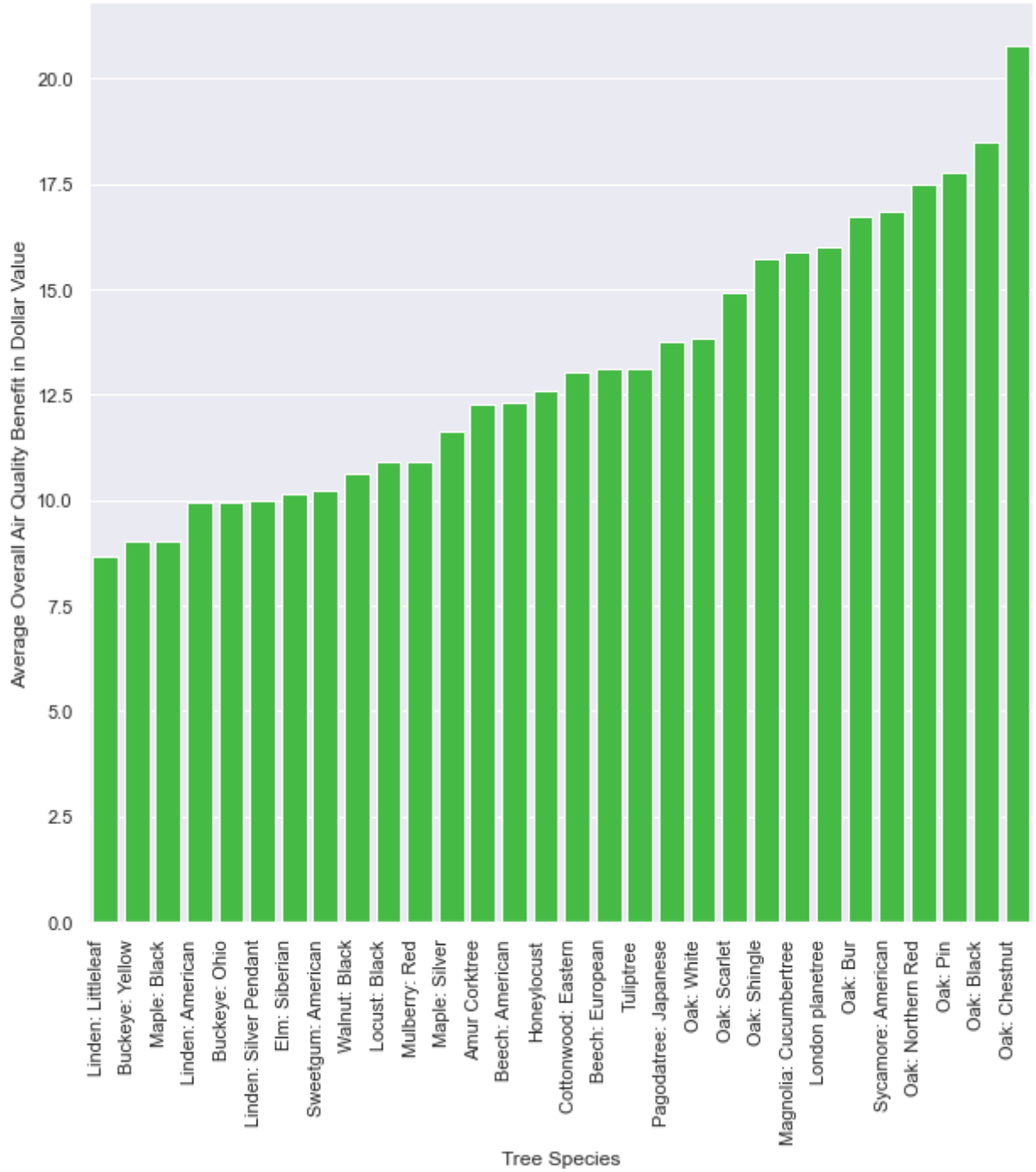

Average Energy (Gas) Benefit in Dollar Value
0
20
40
60
80
Tree-of-Heaven
Ash: spp.
Buckeye: Yellow
Linden: American
Sweetgum: American
Buckeye: Ohio
Elm: Siberian
Walnut: Black
Oak: White
Linden: Silver Pendant
Beech: American
Ash: White
Beech: European
Maple: Silver
Amur Corktree
Mulberry: Red
Oak: Scarlet
Honeylocust
Oak: Bur
Tuliptree
Cottonwood: Eastern
Locust: Black
London planetree
Oak: Shingle
Magnolia: Cucumbertree
Oak: Pin
Sycamore: American
Oak: Northern Red
Oak: Black
Oak: Chestnut
Tree Species

Oak trees provide the most energy-related benefits.

Top 30 tree species that provide the most overall air quality benefits

In [57]:
```
air_quality_total_benefit = tree_stat.sort_values([("air_quality_benfits_total_dollar_value","mean")])
air_quality_total_benefit = air_quality_total_benefit.tail(30)
```

In [58]:
```
air_quality_total_benefit_bar_plot = sns.barplot(x=air_quality_total_benefit["common_name"], y=air_quali
air_quality_total_benefit_bar_plot.set_xticklabels(air_quality_total_benefit_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
air_quality_total_benefit_bar_plot.set_xlabel("Tree Species", fontsize = 12)
air_quality_total_benefit_bar_plot.set_ylabel("Average Overall Air Quality Benefit in Dollar Value", for
plt.gcf().set_size_inches(10,10)
```



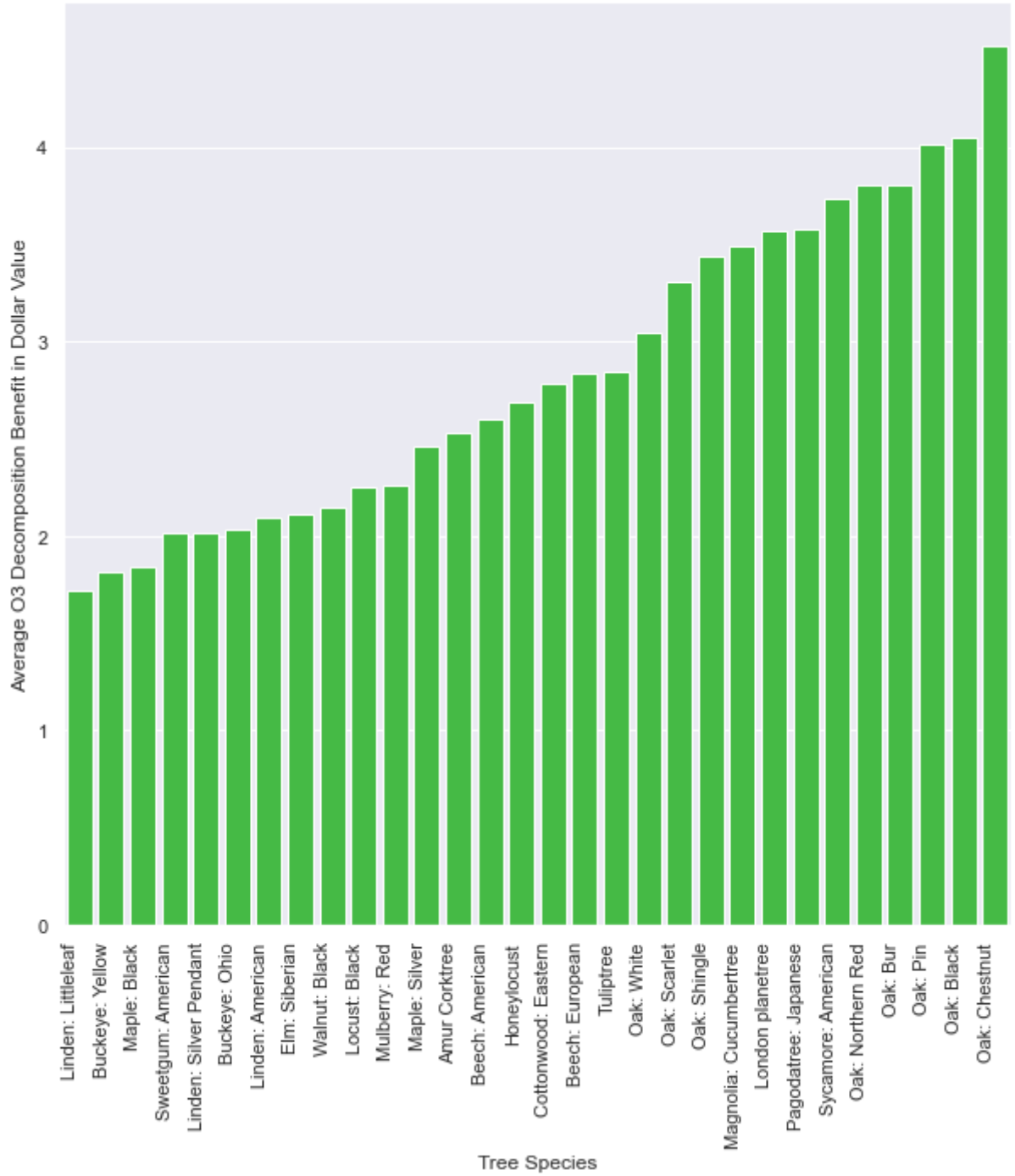
There are many pollutants that can impact air quality, and trees provide values by decomposing or absorbing these chemicals. These

There are many pollutants that can impact air quality, and trees provide values by decomposing or absorbing these chemicals. These pollutants include O3 (ozone), SO2(sulfur dioxide), NO2(nitrogen dioxide), and PM10. We can explore which trees provide the most values in terms of reducing these pollutants.

O3 benefits

In [59]:
```
o3_air_quality_total_benefit = tree_stat.sort_values([("air_quality_benfits_o3dep_dollar_value","mean")]
o3_air_quality_total_benefit = o3_air_quality_total_benefit.tail(30)
```

In [60]:
```
o3_air_quality_total_benefit = sns.barplot(x=o3_air_quality_total_benefit["common_name"], y=o3_air_quali
o3_air_quality_total_benefit.set_xticklabels(o3_air_quality_total_benefit.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
o3_air_quality_total_benefit.set_xlabel("Tree Species", fontsize = 12)
o3_air_quality_total_benefit.set_ylabel("Average O3 Decomposition Benefit in Dollar Value", fontsize = 1
plt.gcf().set_size_inches(10,10)
```

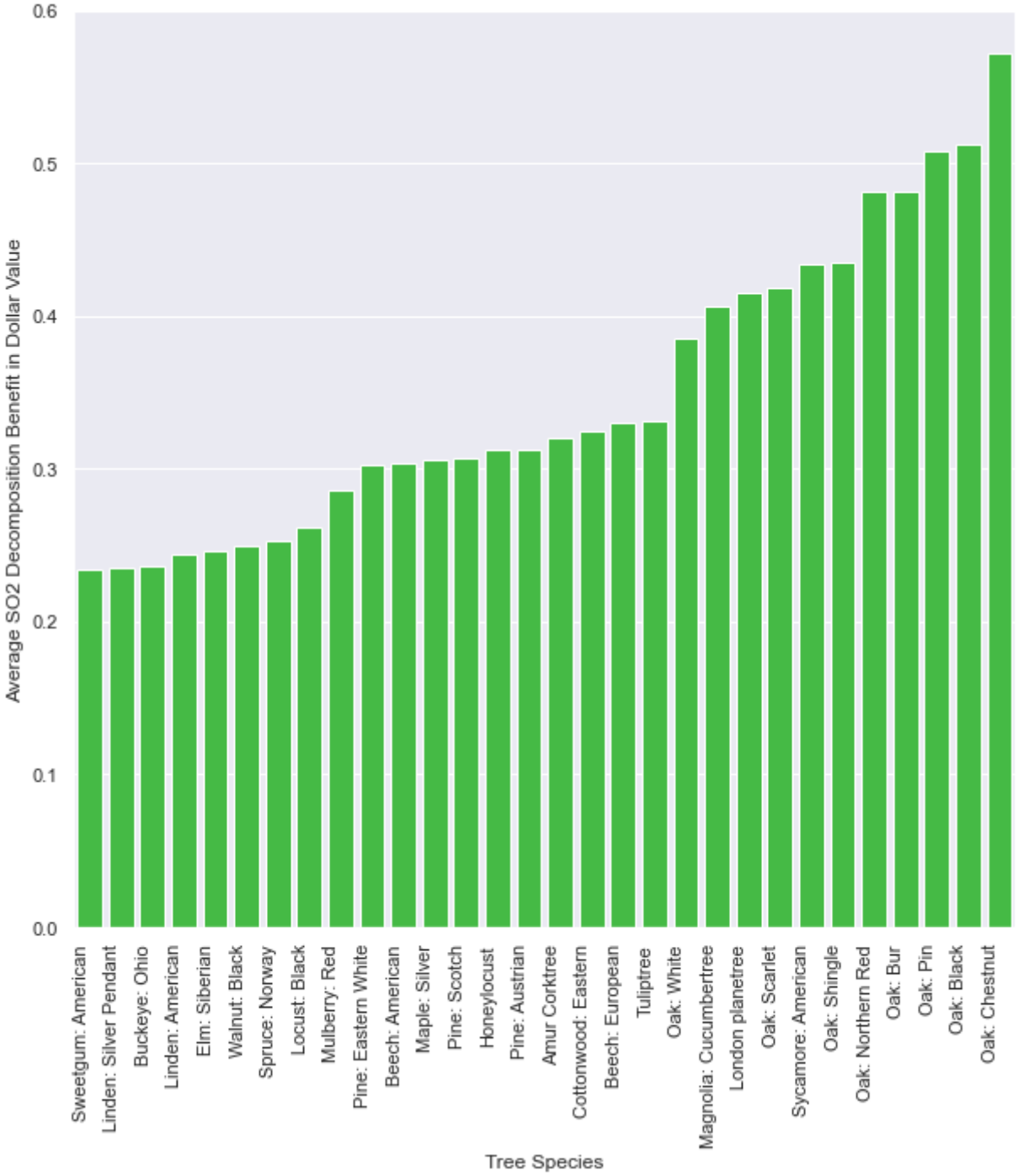


Oak trees absorb the most O3 and offer the most benefit in dollar values.

SO2 Benefits

In [61]:
```
so2_air_quality_total_benefit = tree_stat.sort_values([("air_quality_benfits_so2dep_dollar_value","mean"
so2_air_quality_total_benefit = so2_air_quality_total_benefit.tail(30)
```

In [62]:
```
so2_air_quality_total_benefit = sns.barplot(x=so2_air_quality_total_benefit["common_name"], y=so2_air_qu
so2_air_quality_total_benefit.set_xticklabels(so2_air_quality_total_benefit.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
so2_air_quality_total_benefit.set_xlabel("Tree Species", fontsize = 12)
so2_air_quality_total_benefit.set_ylabel("Average SO2 Decomposition Benefit in Dollar Value", fontsize =
plt.gcf().set_size_inches(10,10)
```

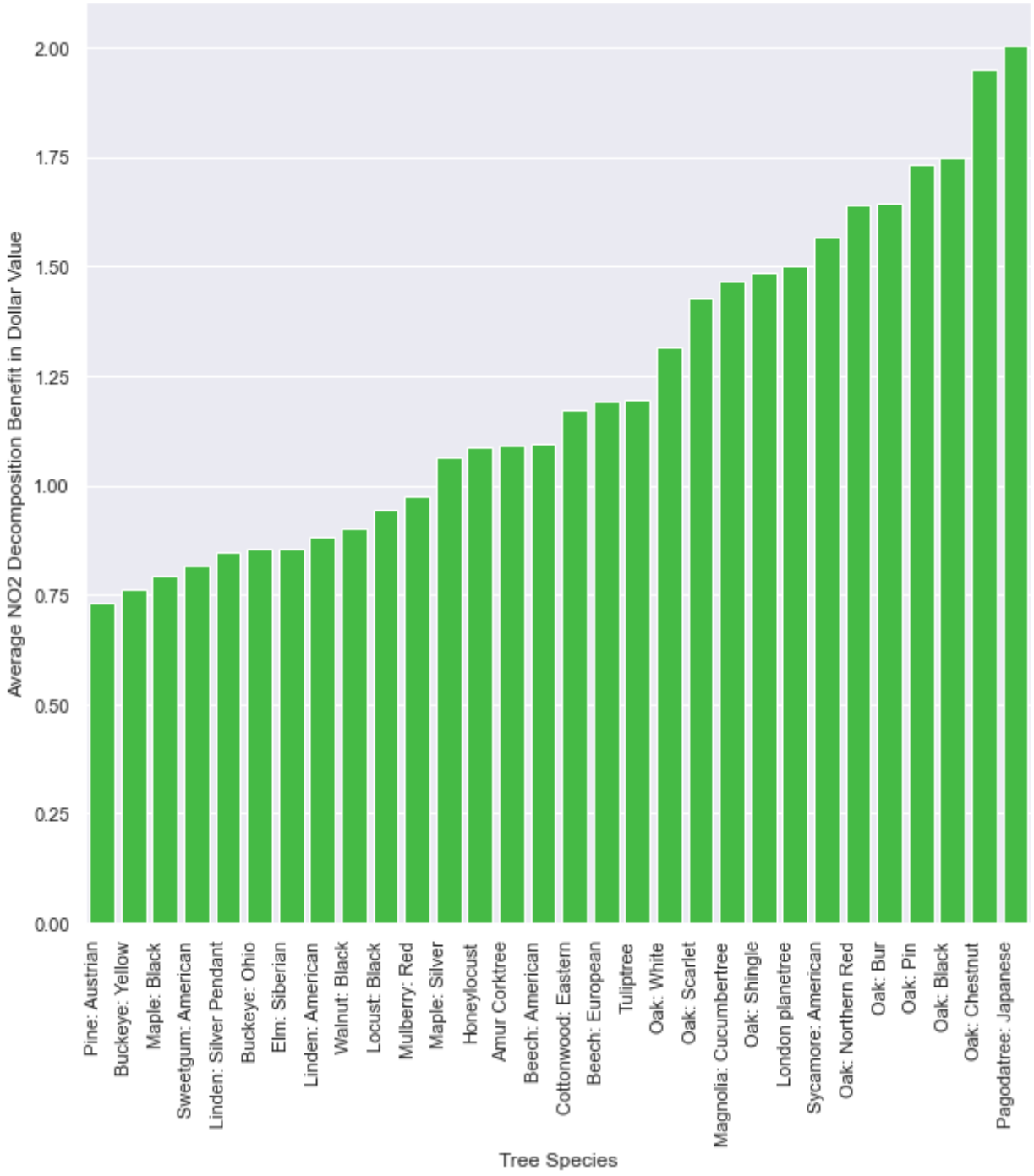

Average SO2 Decomposition Benefit in Dollar Value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Sweetgum: American
Linden: Silver Pendant
Buckeye: Ohio
Linden: American
Elm: Siberian
Walnut: Black
Spruce: Norway
Locust: Black
Mulberry: Red
Pine: Eastern White
Beech: American
Maple: Silver
Pine: Scotch
Honeylocust
Pine: Austrian
Amur Corktree
Cottonwood: Eastern
Beech: European
Tuliptree
Oak: White
Magnolia: Cucumbertree
London planetree
Oak: Scarlet
Sycamore: American
Oak: Shingle
Oak: Northern Red
Oak: Bur
Oak: Pin
Oak: Black
Oak: Chestnut
Tree Species

Oak trees absorb the most SO2 and offer the most benefit in dollar values.

NO2 Benefits

In [63]:
```
no2_air_quality_total_benefit = tree_stat.sort_values([("air_quality_benfits_no2dep_dollar_value","mean"
no2_air_quality_total_benefit = no2_air_quality_total_benefit.tail(30)
```

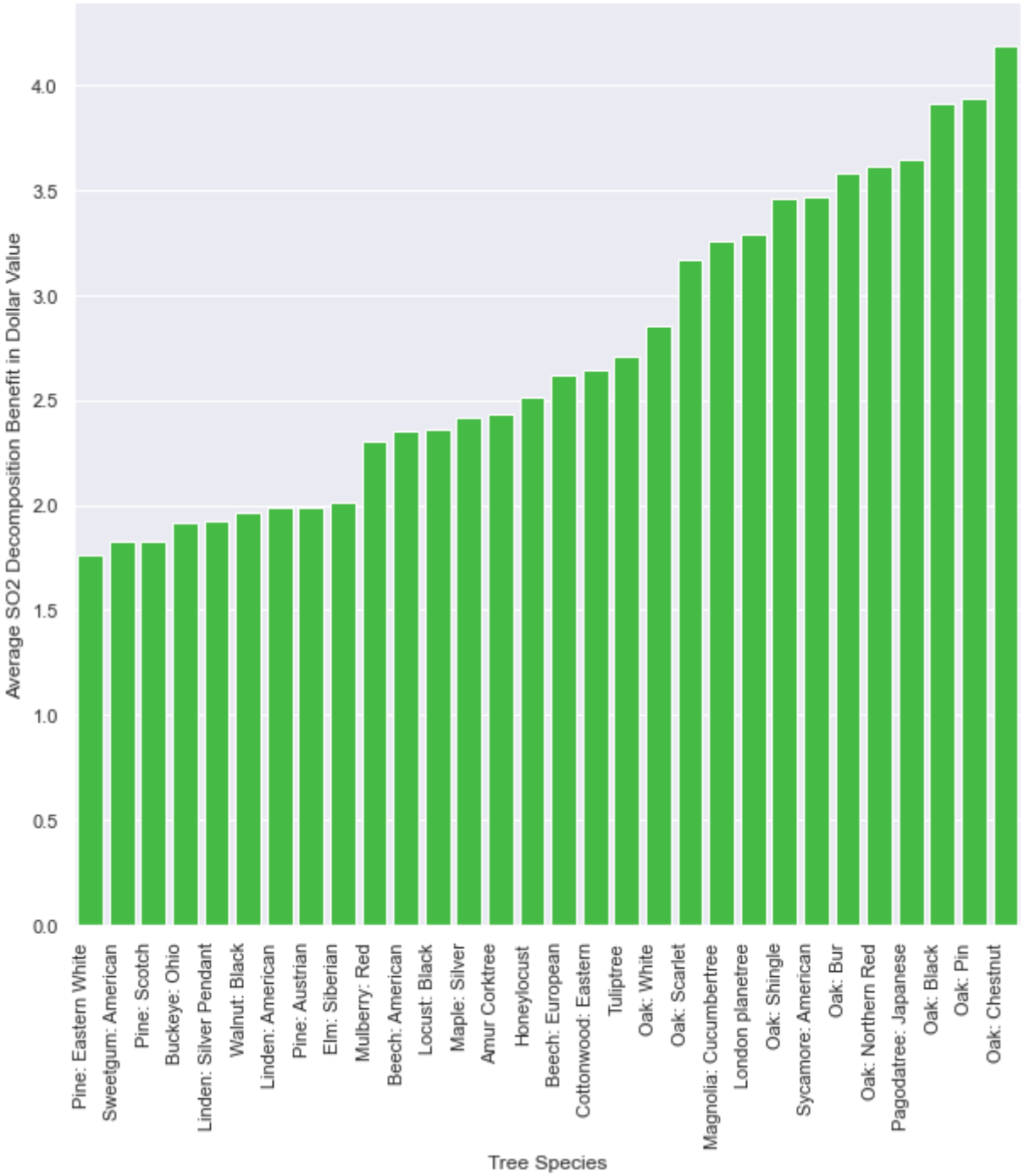
In [64]:
```
no2_air_quality_total_benefit = sns.barplot(x=no2_air_quality_total_benefit["common_name"], y=no2_air_qu
no2_air_quality_total_benefit.set_xticklabels(no2_air_quality_total_benefit.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
no2_air_quality_total_benefit.set_xlabel("Tree Species", fontsize = 12)
no2_air_quality_total_benefit.set_ylabel("Average NO2 Decomposition Benefit in Dollar Value", fontsize =
plt.gcf().set_size_inches(10,10)
```



Surprisingly, for NO2, Japanese Pegoda trees absorb and decompose the most and offer the most benefit in dollar values.

In [65]:
```
pm10_air_quality_total_benefit = tree_stat.sort_values([("air_quality_benfits_pm10depdollar_value","mean
pm10_air_quality_total_benefit = pm10_air_quality_total_benefit.tail(30)
```

In [66]:
```
pm10_air_quality_total_benefit = sns.barplot(x=pm10_air_quality_total_benefit["common_name"], y=pm10_air
pm10_air_quality_total_benefit.set_xticklabels(pm10_air_quality_total_benefit.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
pm10_air_quality_total_benefit.set_xlabel("Tree Species", fontsize = 12)
pm10_air_quality_total_benefit.set_ylabel("Average SO2 Decomposition Benefit in Dollar Value", fontsize
plt.gcf().set_size_inches(10,10)
```

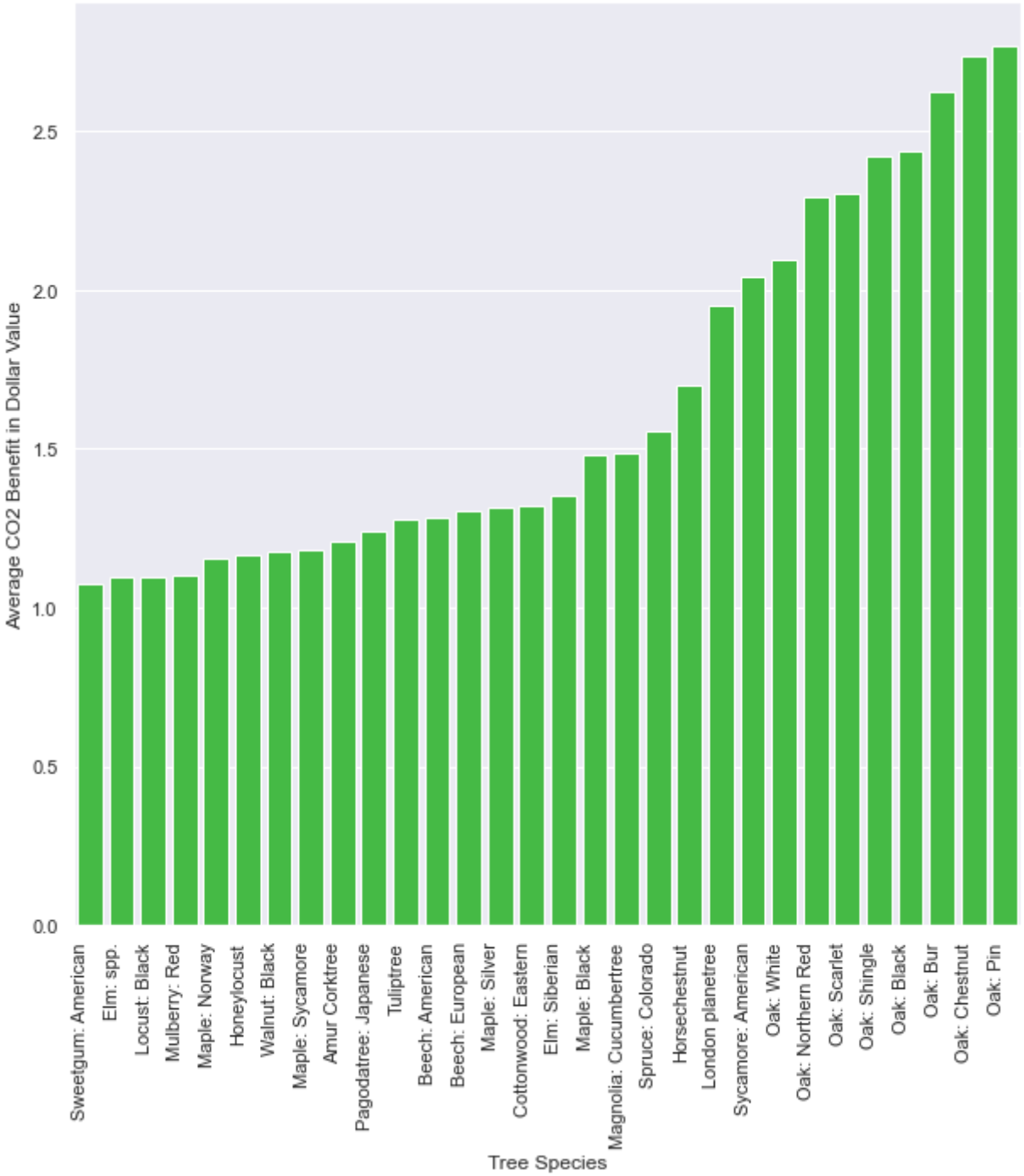

Average SO2 Decomposition Benefit in Dollar Value
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Pine: Eastern White
Sweetgum: American
Pine: Scotch
Buckeye: Ohio
Linden: Silver Pendant
Walnut: Black
Linden: American
Pine: Austrian
Elm: Siberian
Mulberry: Red
Beech: American
Locust: Black
Maple: Silver
Amur Corktree
Honeylocust
Beech: European
Cottonwood: Eastern
Tuliptree
Oak: White
Oak: Scarlet
Magnolia: Cucumbertree
London planetree
Oak: Shingle
Sycamore: American
Oak: Bur
Oak: Northern Red
Pagodatree: Japanese
Oak: Black
Oak: Pin
Oak: Chestnut
Tree Species

Once again oak trees provide the most benefits in terms of PM10 reduction.

Top 30 tree species that provide the most CO2 related benefits (CO2 sequester, CO2 reduction, etc.)

In [67]:
```
co2_total_benefit = tree_stat.sort_values([("co2_benefits_dollar_value","mean")])
co2_total_benefit = co2_total_benefit.tail(30)
```

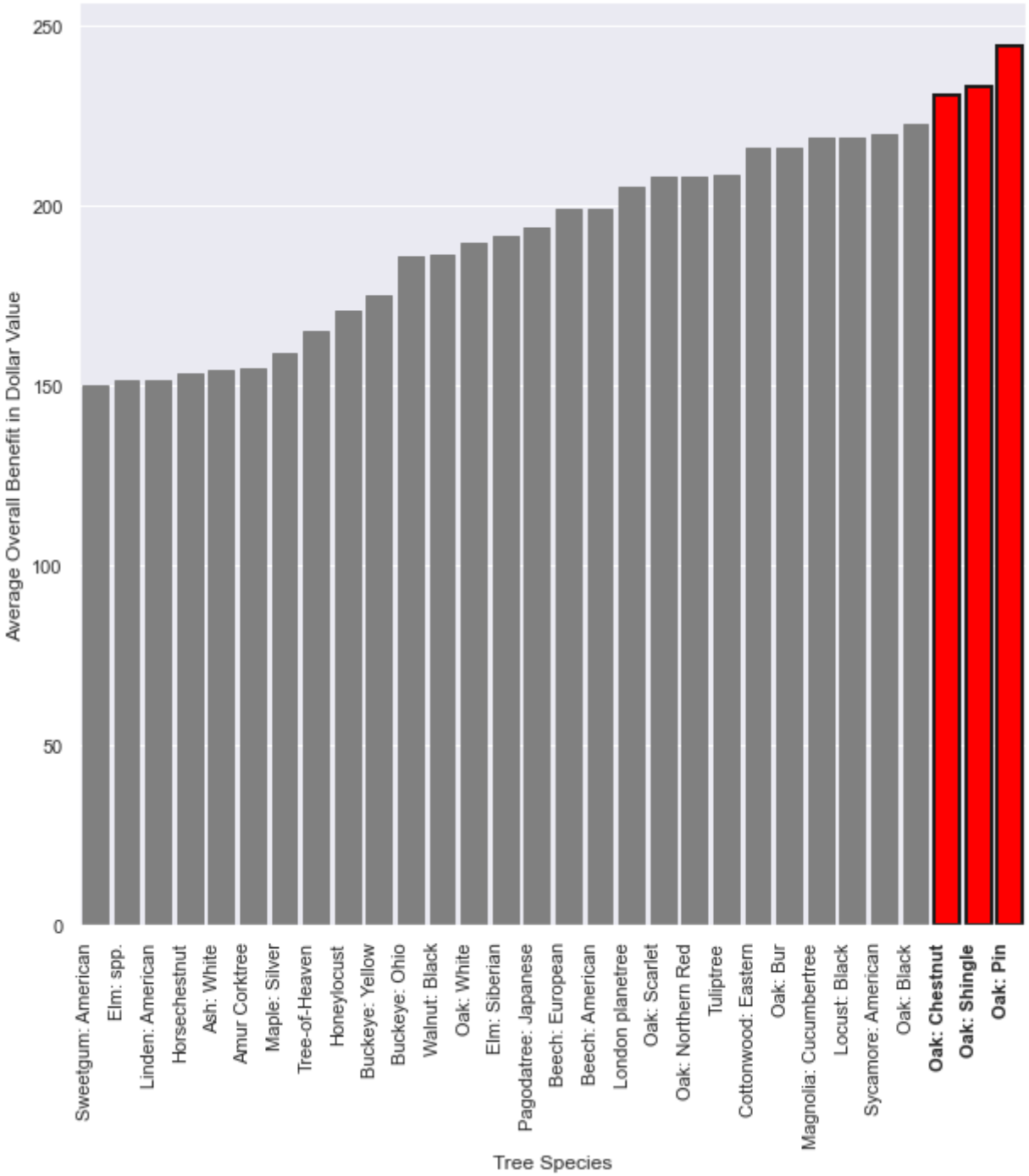
In [68]:
```
co2_total_benefit_bar_plot = sns.barplot(x=co2_total_benefit["common_name"], y=co2_total_benefit["co2_be
co2_total_benefit_bar_plot.set_xticklabels(co2_total_benefit_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
co2_total_benefit_bar_plot.set_xlabel("Tree Species", fontsize = 12)
co2_total_benefit_bar_plot.set_ylabel("Average CO2 Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```


Average CO2 Benefit in Dollar Value
0.0
0.5
1.0
1.5
2.0
2.5
Sweetgum: American
Elm: spp.
Locust: Black
Mulberry: Red
Maple: Norway
Honeylocust
Walnut: Black
Maple: Sycamore
Amur Corktree
Pagodatree: Japanese
Tuliptree
Beech: American
Beech: European
Maple: Silver
Cottonwood: Eastern
Elm: Siberian
Maple: Black
Magnolia: Cucumbertree
Spruce: Colorado
Horsechestnut
London planetree
Sycamore: American
Oak: White
Oak: Northern Red
Oak: Scarlet
Oak: Shingle
Oak: Black
Oak: Bur
Oak: Chestnut
Oak: Pin
Tree Species

Top 30 tree species that provide overall the most benefits, with top 3 highlighted

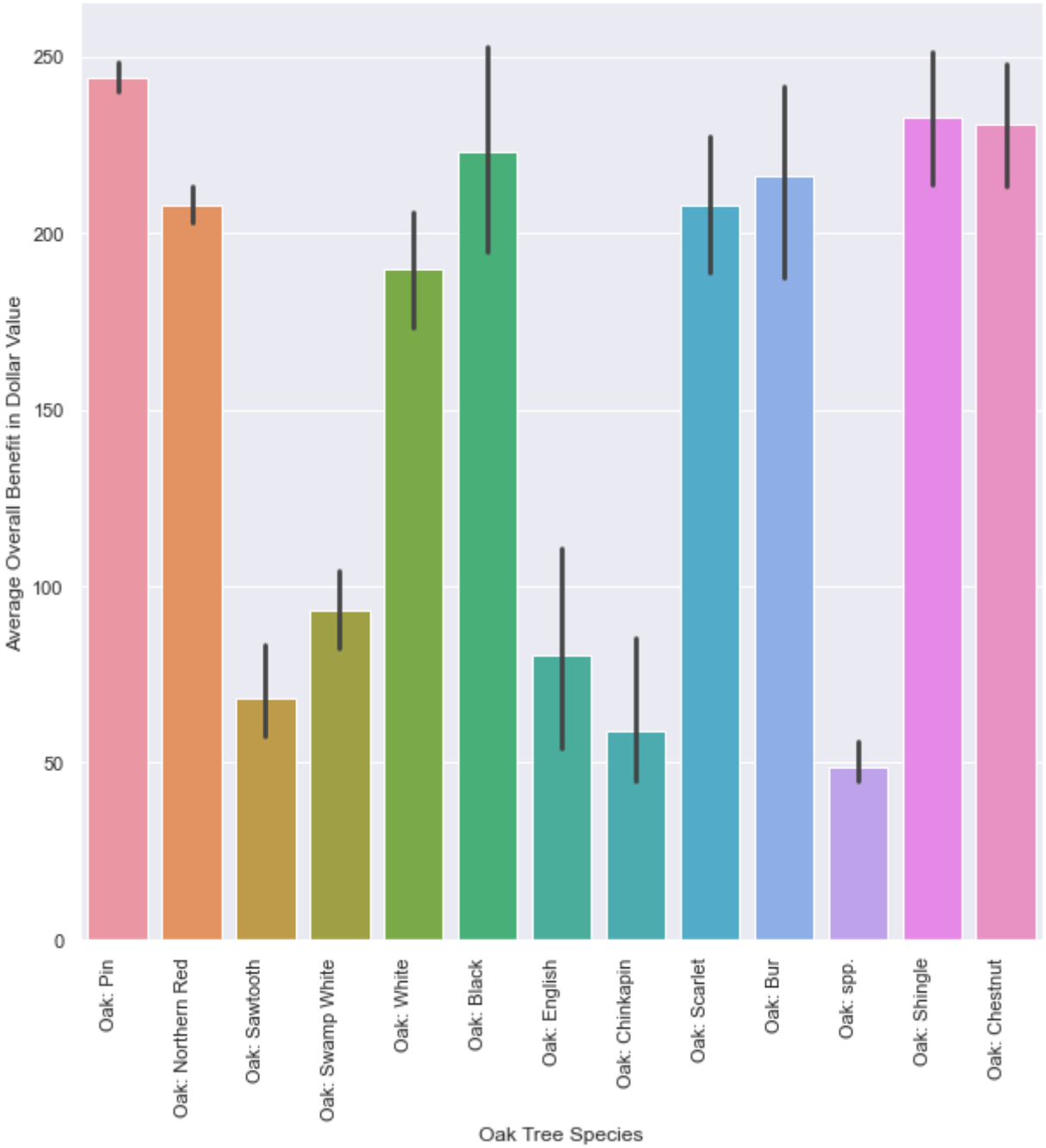
In [69]:
```
overall_total_benefit = tree_stat.sort_values([("overall_benefits_dollar_value","mean")])
overall_total_benefit = overall_total_benefit.tail(30)
```

In [70]:
```
overall_total_benefit_bar_plot = sns.barplot(x=overall_total_benefit["common_name"], y=overall_total_ben
overall_total_benefit_bar_plot.set_xticklabels(overall_total_benefit_bar_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
overall_total_benefit_bar_plot.set_xlabel("Tree Species", fontsize = 12)
overall_total_benefit_bar_plot.set_ylabel("Average Overall Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
for bar in overall_total_benefit_bar_plot.patches:
    if bar.get_height() > 225:
        bar.set_color('red')
    else:
        bar.set_color('grey')

for i,t in enumerate(overall_total_benefit_bar_plot.get_xticklabels()):
    if t.get_text() in ["Oak: Pin", "Oak: Shingle", "Oak: Chestnut"]:
        ## bold ticklabels
        t.set_weight("bold")
        ## bar edges
        overall_total_benefit_bar_plot.patches[i].set_edgecolor("k")
        overall_total_benefit_bar_plot.patches[i].set_linewidth(2)
```


250
200
150
100
50
0
Average Overall Benefit in Dollar Value
Sweetgum: American
Elm: spp.
Linden: American
Horsechestnut
Ash: White
Amur Corktree
Maple: Silver
Tree-of-Heaven
Honeylocust
Buckeye: Yellow
Buckeye: Ohio
Walnut: Black
Oak: White
Elm: Siberian
Pagodatree: Japanese
Beech: European
Beech: American
London planetree
Oak: Scarlet
Oak: Northern Red
Tuliptree
Cottonwood: Eastern
Oak: Bur
Magnolia: Cucumbertree
Locust: Black
Sycamore: American
Oak: Black
Oak: Chestnut
Oak: Shingle
Oak: Pin
Tree Species

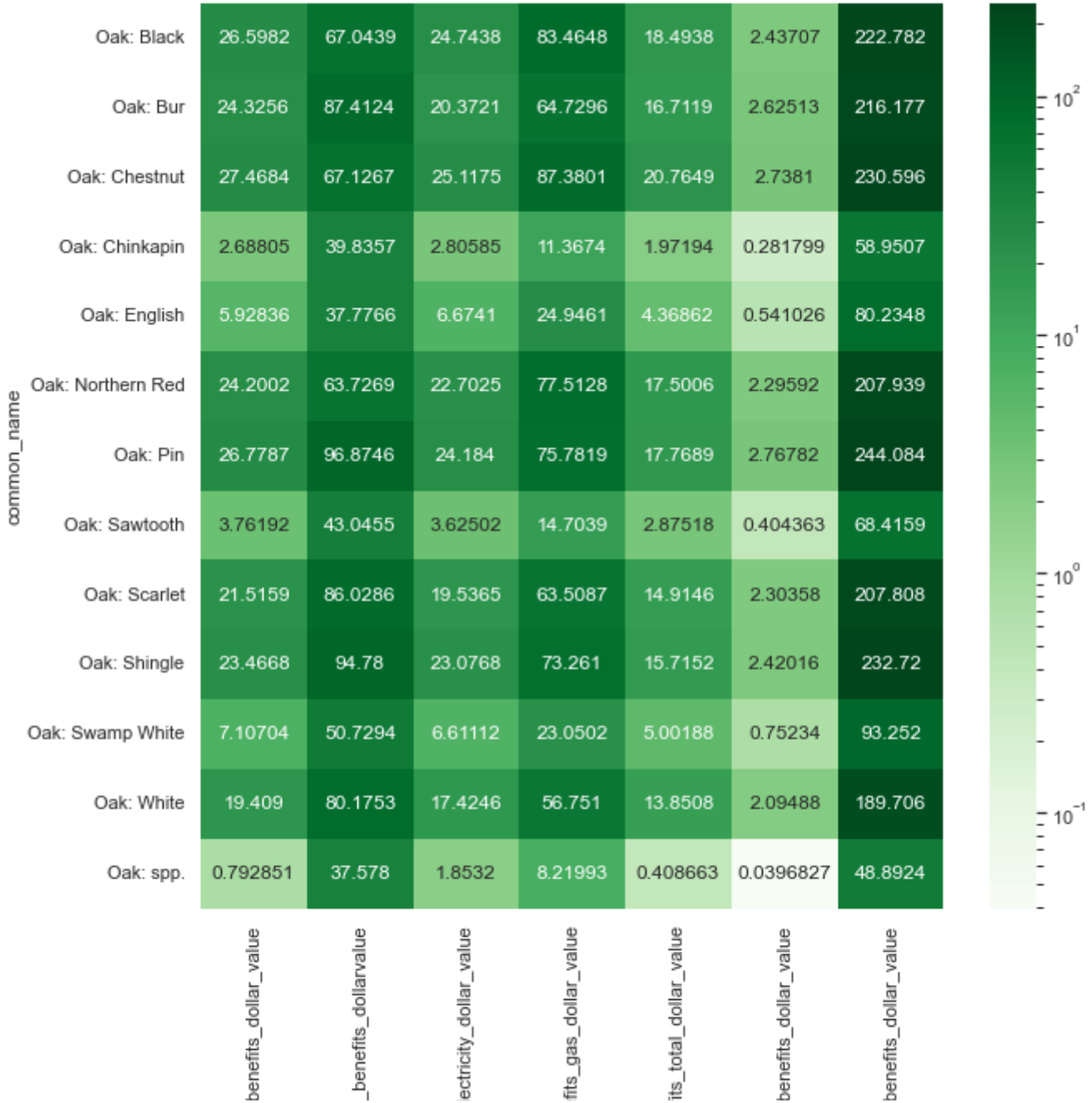
Based on the exploration, it seems like different kinds of oak trees offer overall the most benefits. They also offer great benefits in the other categories explored too. Let's compare the different oak species more in-depth.

In [71]:
```
oaks = df_trees[df_trees['common_name'].str.contains('Oak')]

#Again only take oaks species that contains more than 10 data points.
oaks_count = oaks.groupby("common_name").agg("count").reset_index()
oaks_count = oaks_count[oaks_count["id"] >= 10]
oaks_selected = oaks_count["common_name"].unique()

oaks_to_plot = oaks[oaks.common_name.isin(oaks_selected)]
oaks_bar_plot = sns.barplot(x=oaks_to_plot["common_name"], y=oaks_to_plot["overall_benefits_dollar_value
oaks_bar_plot.set_xticklabels(oaks_bar_plot.get_xticklabels(),
                              rotation=90,
                              horizontalalignment='right')
oaks_bar_plot.set_xlabel("Oak Tree Species", fontsize = 12)
oaks_bar_plot.set_ylabel("Average Overall Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```



There is big variance in benefits between different oak species. We can build a heatmap to more easily notice the differences.

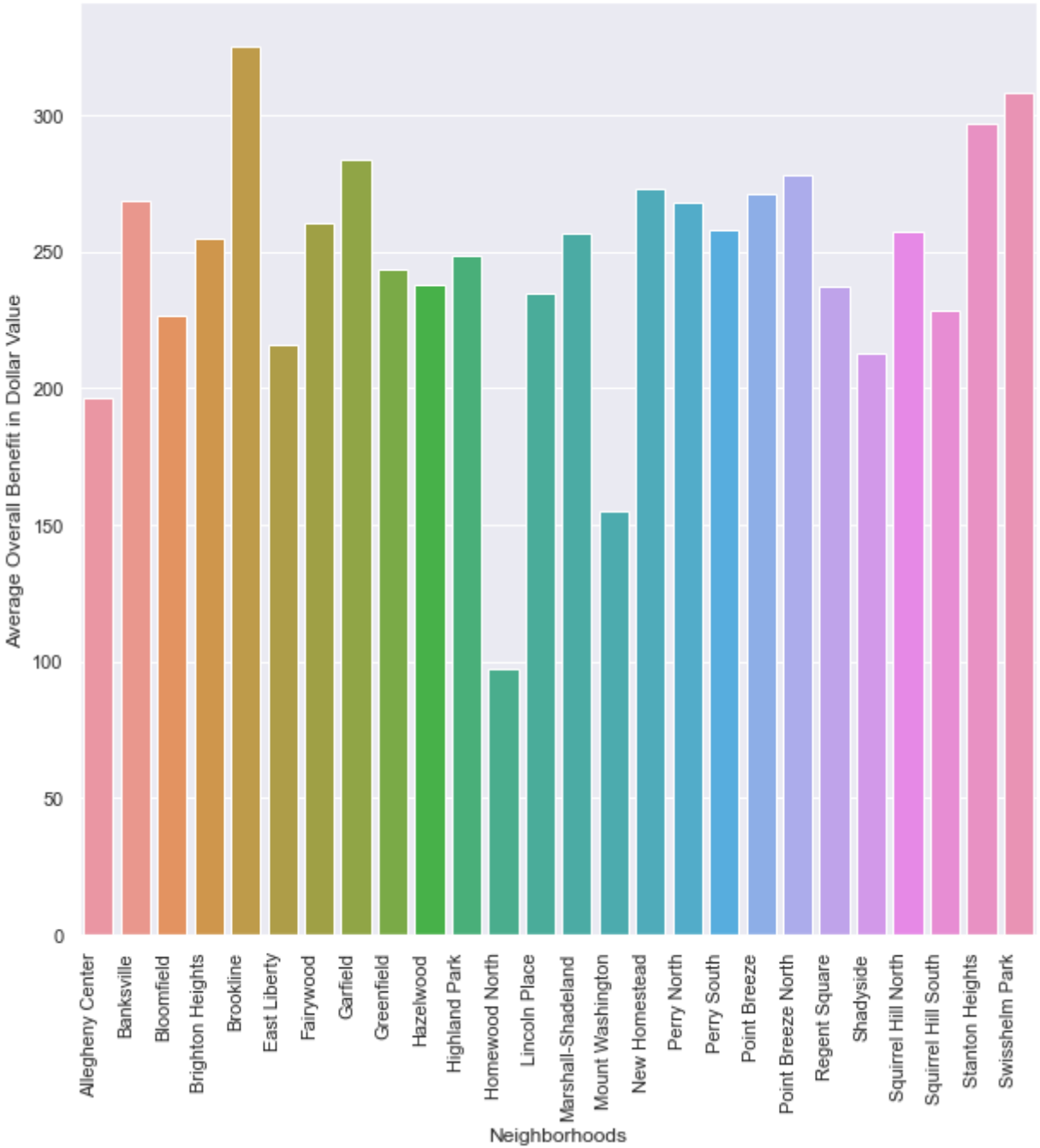
In [72]:
```
oaks_group = oaks_to_plot.groupby("common_name").agg('mean')
grouped_oaks_plot = sns.heatmap(oaks_group[["stormwater_benefits_dollar_value","property_value_benefits_
plt.gcf().set_size_inches(10,10)
```



It is clear that some benefits, such as property value benefit, will be dependent on the neighborhood since the property prices can differ drastically. However, now we wonder if a tree will provide different values in different neighborhoods for things that should not be really neighborhood-dependent, such as energy-saving or air quality benefit. To investigate this, I picked the "Oak: Pin" tree as the candidate to graph because this tree species provided overall the most benefit, as shown by explorations done above.

In [73]:
```
oak_tree_neighborhood = df_trees[df_trees["common_name"] == "Oak: Pin"]
oak_tree_neighborhood = oak_tree_neighborhood.groupby("neighborhood").agg(["mean", "count"])

oak_tree_neighborhood = oak_tree_neighborhood[oak_tree_neighborhood["id"]["count"] >= 10]
oak_tree_neighborhood = oak_tree_neighborhood.reset_index()
```

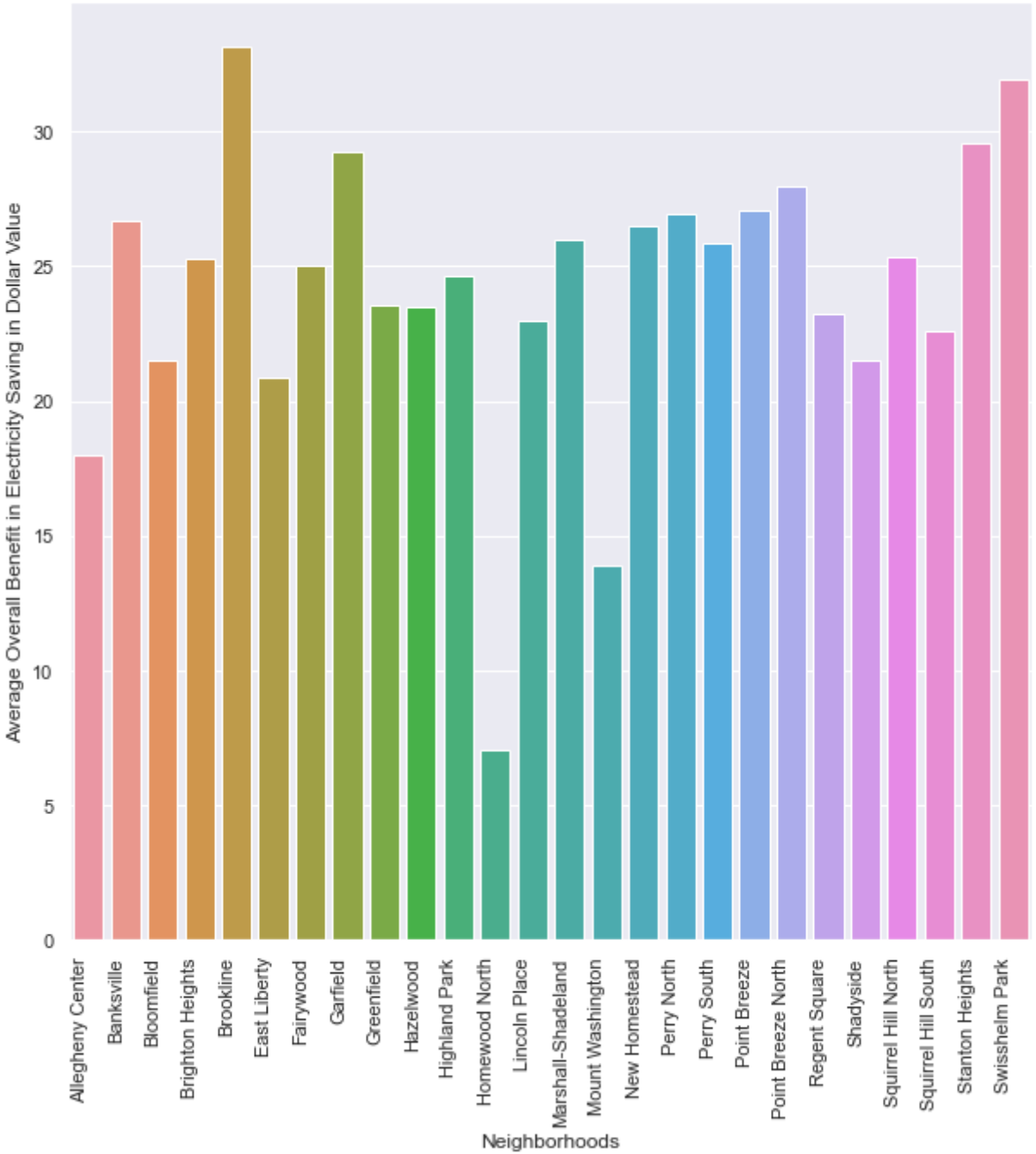
In [74]:
```
neighborhood_diff_plot = sns.barplot(x=oak_tree_neighborhood["neighborhood"], y=oak_tree_neighborhood["c
neighborhood_diff_plot.set_xticklabels(neighborhood_diff_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
neighborhood_diff_plot.set_xlabel("Neighborhoods", fontsize = 12)
neighborhood_diff_plot.set_ylabel("Average Overall Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```



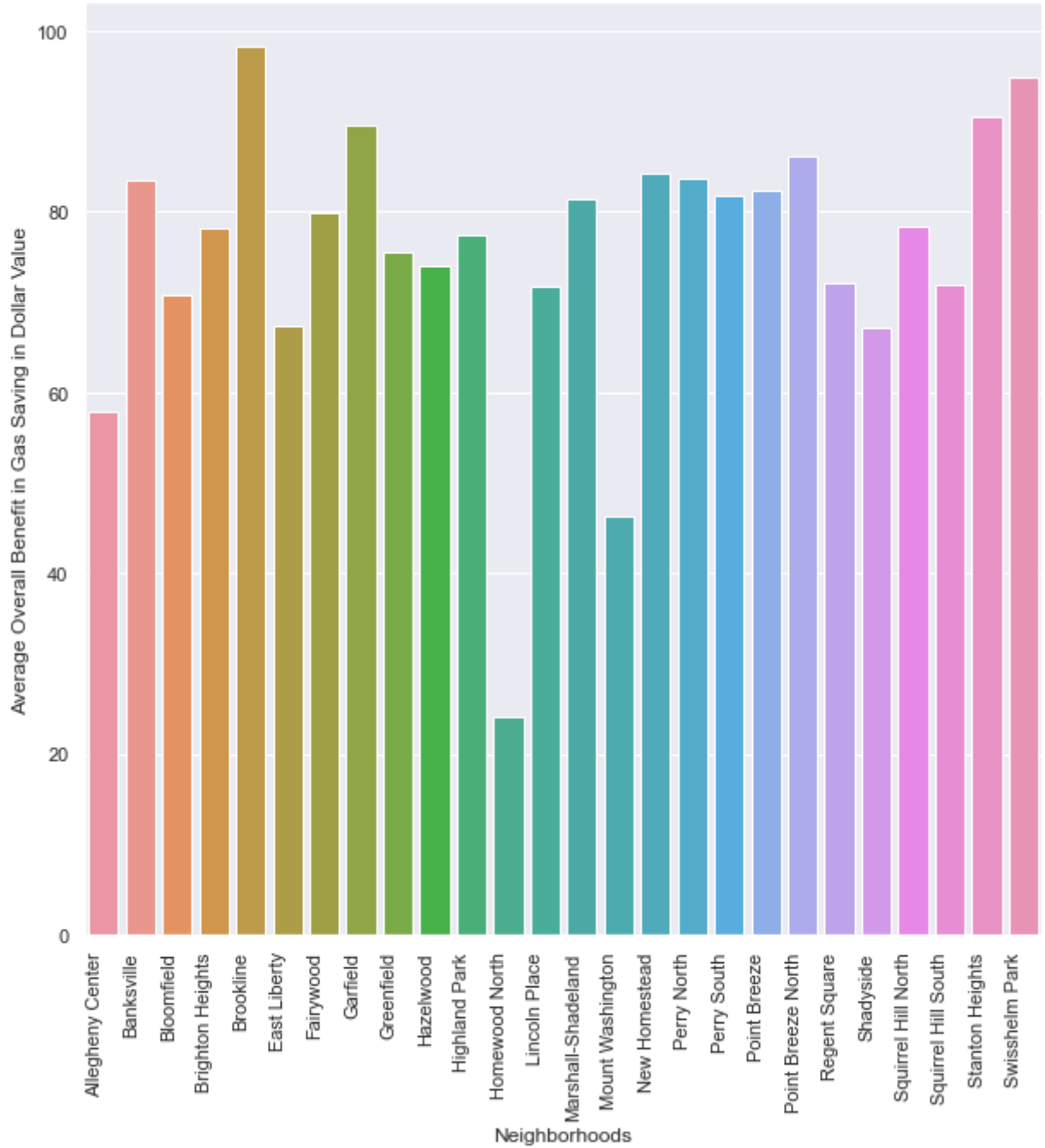
Surprisingly there are big differences between neighborhoods in terms of the overall benefit provided by this tree species. However, this might be purely because of the difference in property benefits, so we need more investigation into some of the sub-categories of overall

benefits.

In [75]:
```
neighborhood_diff_electricity_plot = sns.barplot(x=oak_tree_neighborhood["neighborhood"], y=oak_tree_nei
neighborhood_diff_electricity_plot.set_xticklabels(neighborhood_diff_electricity_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
neighborhood_diff_electricity_plot.set_xlabel("Neighborhoods", fontsize = 12)
neighborhood_diff_electricity_plot.set_ylabel("Average Overall Benefit in Electricity Saving in Dollar V
plt.gcf().set_size_inches(10,10)
```

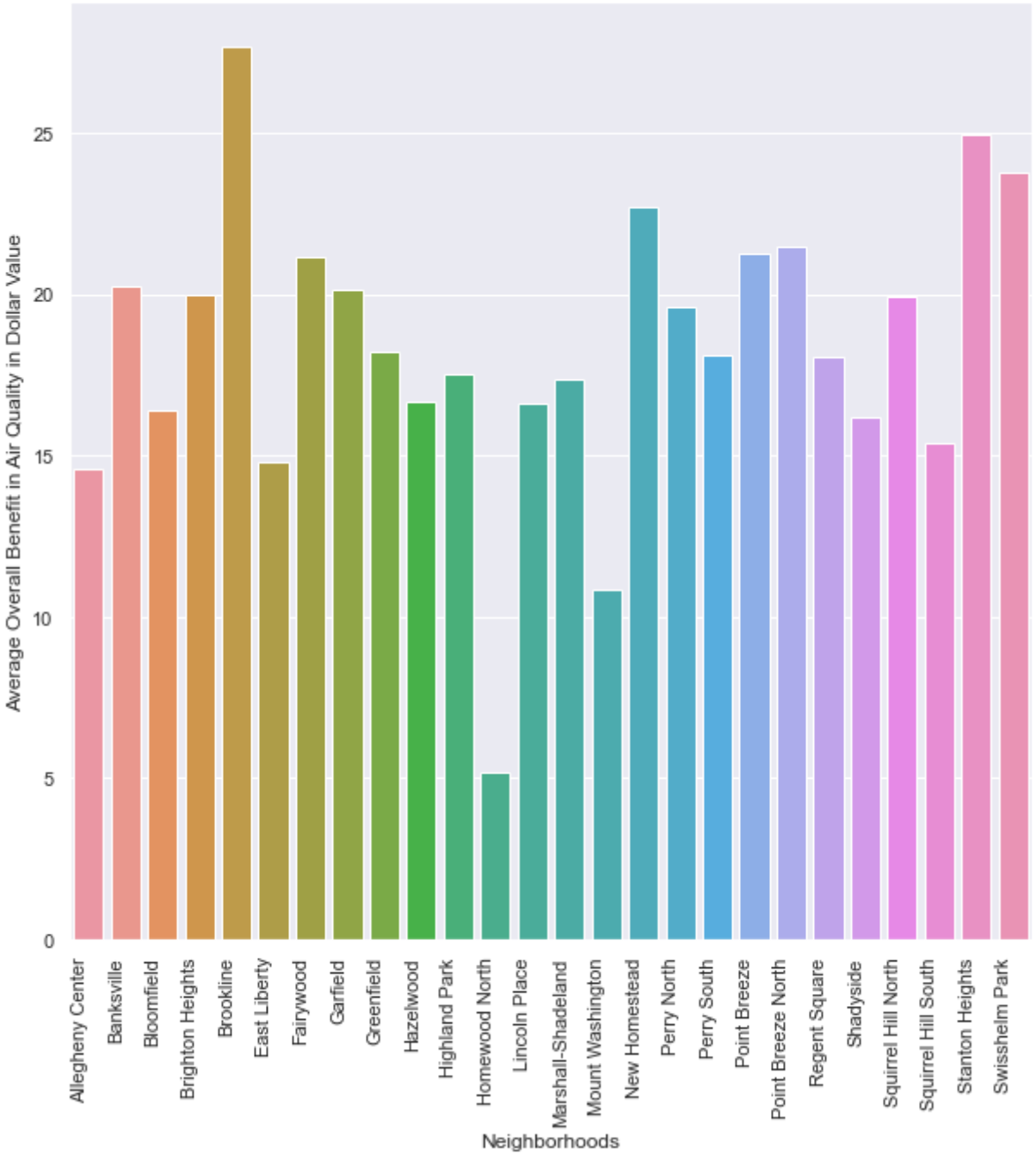

Average Overall Benefit in Electricity Saving in Dollar Value
0
5
10
15
20
25
30
Allegheny Center
Banksville
Bloomfield
Brighton Heights
Brookline
East Liberty
Fairywood
Garfield
Greenfield
Hazelwood
Highland Park
Homewood North
Lincoln Place
Marshall-Shadeland
Mount Washington
New Homestead
Perry North
Perry South
Point Breeze
Point Breeze North
Regent Square
Shadyside
Squirrel Hill North
Squirrel Hill South
Stanton Heights
Swisshelm Park
Neighborhoods

In [76]:
```
neighborhood_diff_gas_plot = sns.barplot(x=oak_tree_neighborhood["neighborhood"], y=oak_tree_neighborhoo
neighborhood_diff_gas_plot.set_xticklabels(neighborhood_diff_gas_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
neighborhood_diff_gas_plot.set_xlabel("Neighborhoods", fontsize = 12)
neighborhood_diff_gas_plot.set_ylabel("Average Overall Benefit in Gas Saving in Dollar Value", fontsize
plt.gcf().set_size_inches(10,10)
```

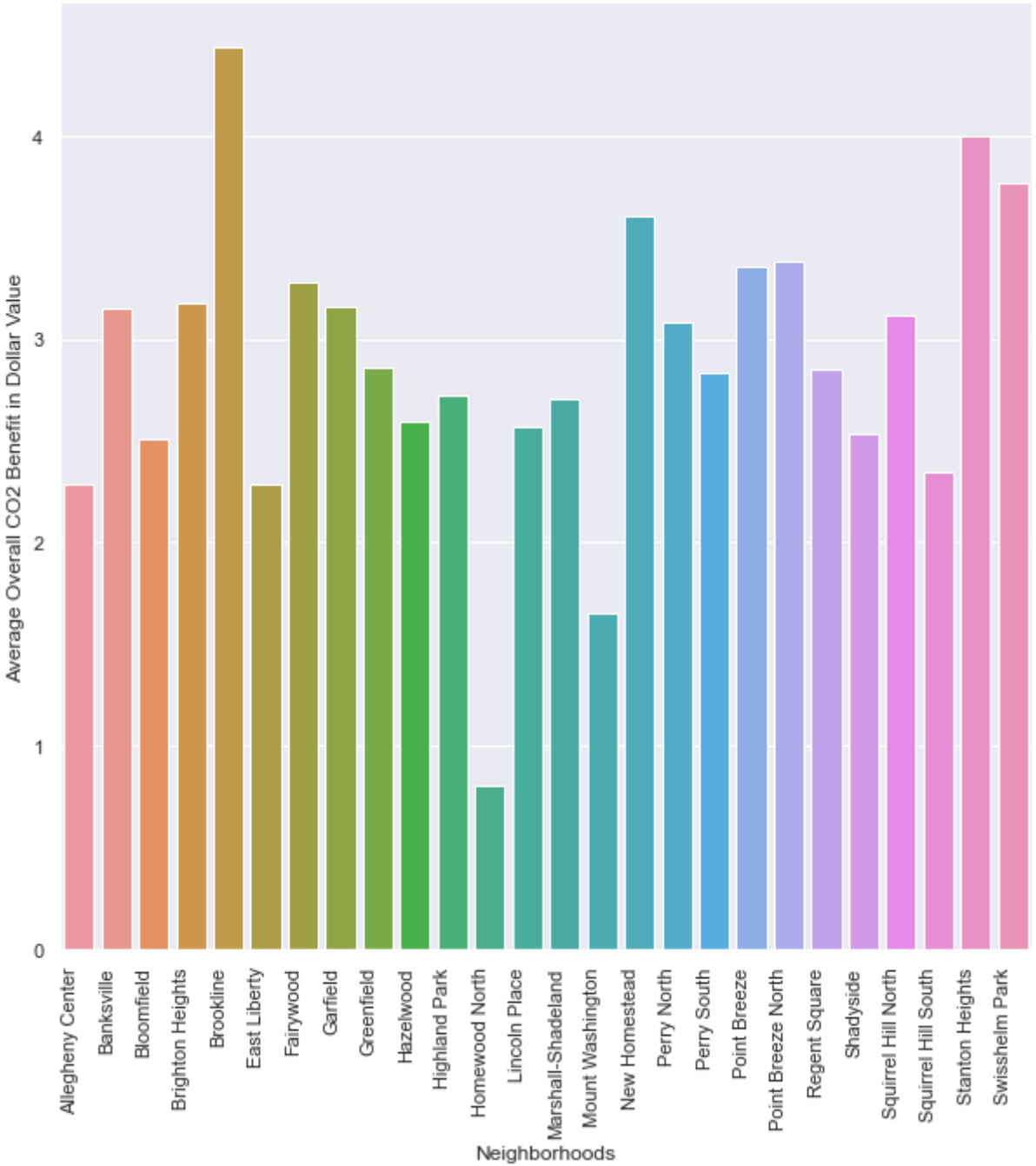


There are big variance in energy saving benefits acorss neighborhoods as well.

In [77]:
```
neighborhood_diff_air_q_plot = sns.barplot(x=oak_tree_neighborhood["neighborhood"], y=oak_tree_neighborh
neighborhood_diff_air_q_plot.set_xticklabels(neighborhood_diff_air_q_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
neighborhood_diff_air_q_plot.set_xlabel("Neighborhoods", fontsize = 12)
neighborhood_diff_air_q_plot.set_ylabel("Average Overall Benefit in Air Quality in Dollar Value", fontsi
plt.gcf().set_size_inches(10,10)
```


Average Overall Benefit in Air Quality in Dollar Value
0
5
10
15
20
25
Allegheny Center
Banksville
Bloomfield
Brighton Heights
Brookline
East Liberty
Fairywood
Garfield
Greenfield
Hazelwood
Highland Park
Homewood North
Lincoln Place
Marshall-Shadeland
Mount Washington
New Homestead
Perry North
Perry South
Point Breeze
Point Breeze North
Regent Square
Shadyside
Squirrel Hill North
Squirrel Hill South
Stanton Heights
Swisshelm Park
Neighborhoods

In [78]:
```
neighborhood_diff_co2_plot = sns.barplot(x=oak_tree_neighborhood["neighborhood"], y=oak_tree_neighborhoo
neighborhood_diff_co2_plot.set_xticklabels(neighborhood_diff_co2_plot.get_xticklabels(),
                           rotation=90,
                           horizontalalignment='right')
neighborhood_diff_co2_plot.set_xlabel("Neighborhoods", fontsize = 12)
neighborhood_diff_co2_plot.set_ylabel("Average Overall CO2 Benefit in Dollar Value", fontsize = 12)
plt.gcf().set_size_inches(10,10)
```



The most surprising finding is that air quality benefits and co2 benefits which should be the least dependent on neighborhoods also show significant variance across the different neighborhoods. Since all of the neighborhoods are located in Pittsburgh, there should not be a

drastic difference in air quality across such a small geographical region over the long term. These two graphs also showed a very similar distribution and a strong correlation.

All of these graphs combined also indicate that some neighborhoods are consistently receiving fewer benefits compared to others, while some are consistently receiving more benefits.

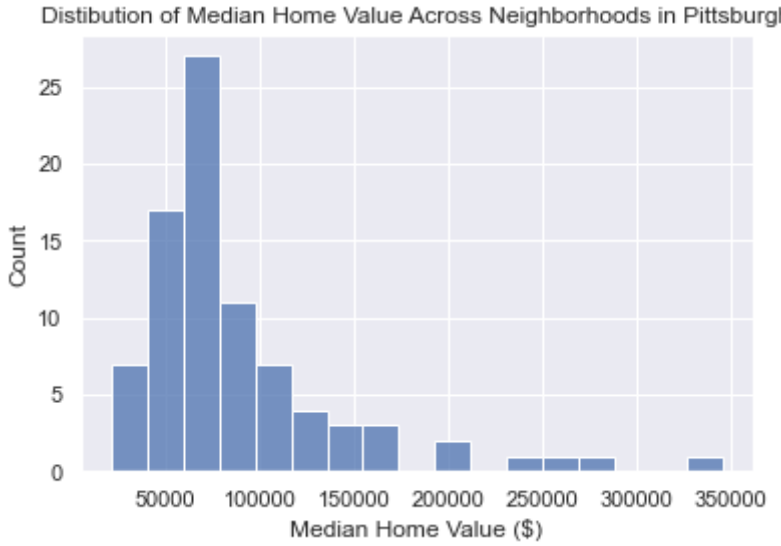
Exploring Trees and Other Neighborhood Factors

In [79]:
```
complete_data = pd.read_csv("cleaned_data/neighborhood_features_data.csv")
```

Median Home Value

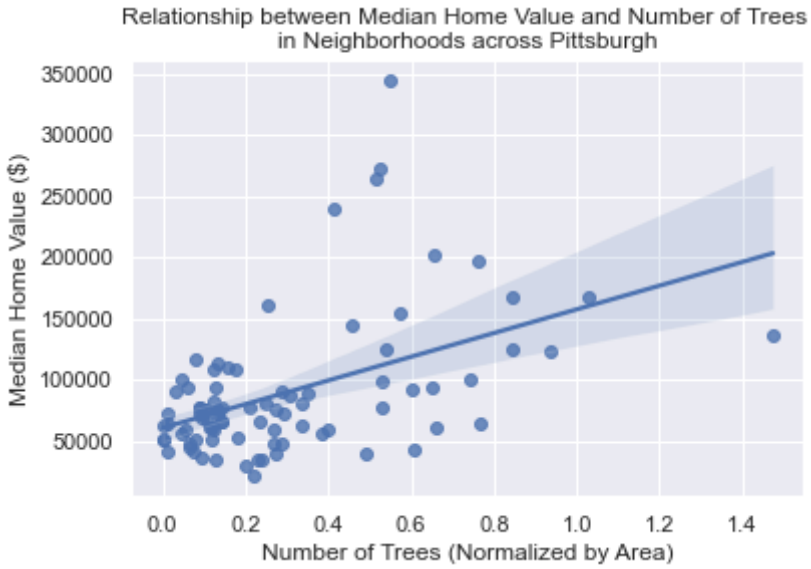
In [80]:
```
home_value_data = complete_data[['median_home_value', 'area_norm_tree_count', 'area_norm_overall_benefit
# remove rows where median_home_value is 0
home_value_data = home_value_data[home_value_data['median_home_value'] != 0]

# plot distribution of median_home_value
plot = sns.histplot(home_value_data['median_home_value'])
plot.set(xlabel = "Median Home Value ($)", title = "Distibution of Median Home Value Across Neighborhooc
```

Out[80]:
```
[Text(0.5, 0, 'Median Home Value ($)'),
 Text(0.5, 1.0, 'Distibution of Median Home Value Across Neighborhoods in Pittsburgh')]
```



In [81]:
```
plot = sns.regplot(x = 'area_norm_tree_count', y = 'median_home_value', data = home_value_data)
plot.set(xlabel = "Number of Trees (Normalized by Area)", ylabel = "Median Home Value ($)",
         title = "Relationship between Median Home Value and Number of Trees \nin Neighborhoods across P
```

Out[81]:
```
[Text(0.5, 0, 'Number of Trees (Normalized by Area)'),
 Text(0, 0.5, 'Median Home Value ($)'),
 Text(0.5, 1.0, 'Relationship between Median Home Value and Number of Trees \nin Neighborhoods across
Pittsburgh')]
```



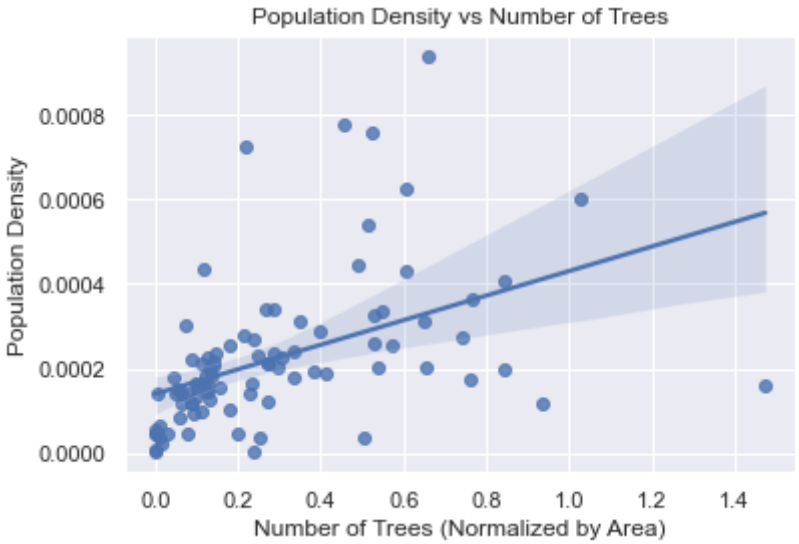
Inference: We observe a slight positive correlation between the median home value and the number of trees. Neighborhoods with higher

numbers of trees tend to have higher Median Home Values on an average.

Population Density

In [82]:
```
plot = sns.regplot(x = 'area_norm_tree_count', y = 'population_density', data = complete_data)
plot.set(xlabel = "Number of Trees (Normalized by Area)", ylabel = "Population Density",
         title = "Population Density vs Number of Trees")
```

Out[82]:
```
[Text(0.5, 0, 'Number of Trees (Normalized by Area)'),
 Text(0, 0.5, 'Population Density'),
 Text(0.5, 1.0, 'Population Density vs Number of Trees')]
```

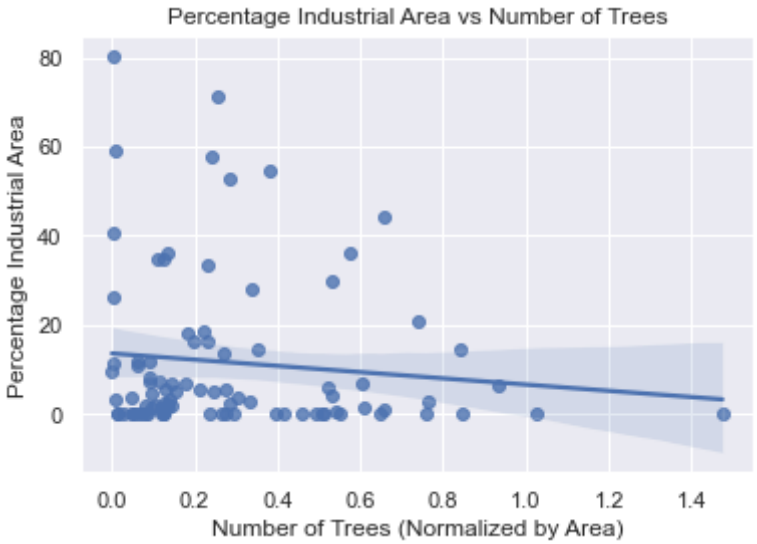


Inference: We also observe a positive correlation between population density and the number of trees, indicating that regions with higher population densities have a larger number of trees per unit area. This seems counter-intuitive given that we would expect areas with larger population densities to have few trees. However, a possible explanation is the presence of regions with sparse vegetation which are not inhabited.

Industrial Areas and Trees

In [83]:
```
plot = sns.regplot(x = 'area_norm_tree_count', y = 'per_industrial_area', data = complete_data)
plot.set(xlabel = "Number of Trees (Normalized by Area)", ylabel = "Percentage Industrial Area",
         title = "Percentage Industrial Area vs Number of Trees")
```

Out[83]:
```
[Text(0.5, 0, 'Number of Trees (Normalized by Area)'),
 Text(0, 0.5, 'Percentage Industrial Area'),
 Text(0.5, 1.0, 'Percentage Industrial Area vs Number of Trees')]
```

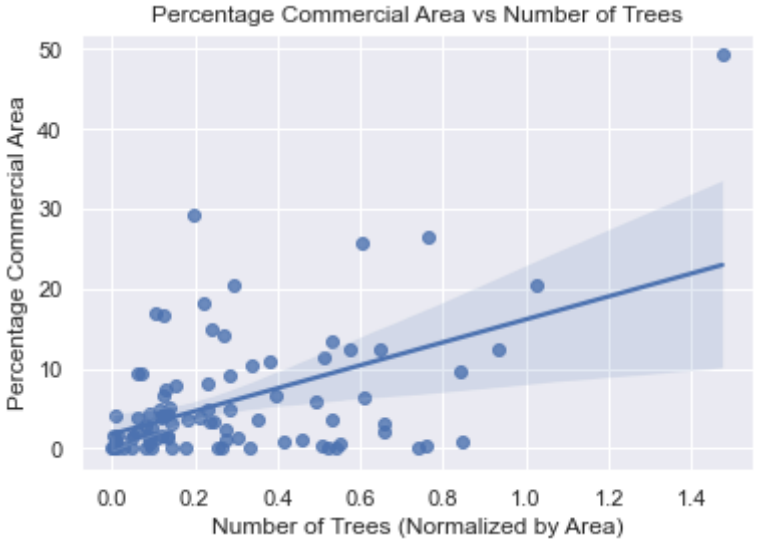


Inference: In this plot, we do not see any trend between the tree density and the percentage of industrial area. However, we do notice that there are no regions with high industrial areas and high tree densities. This is expected since industrialization often corresponds to cutting down of trees or using barren lands.

Commercial Areas and Trees

In [84]:
```
plot = sns.regplot(x = 'area_norm_tree_count', y = 'per_commercial_area', data = complete_data)
plot.set(xlabel = "Number of Trees (Normalized by Area)", ylabel = "Percentage Commercial Area",
         title = "Percentage Commercial Area vs Number of Trees")
```

Out[84]:
```
[Text(0.5, 0, 'Number of Trees (Normalized by Area)'),
 Text(0, 0.5, 'Percentage Commercial Area'),
 Text(0.5, 1.0, 'Percentage Commercial Area vs Number of Trees')]
```

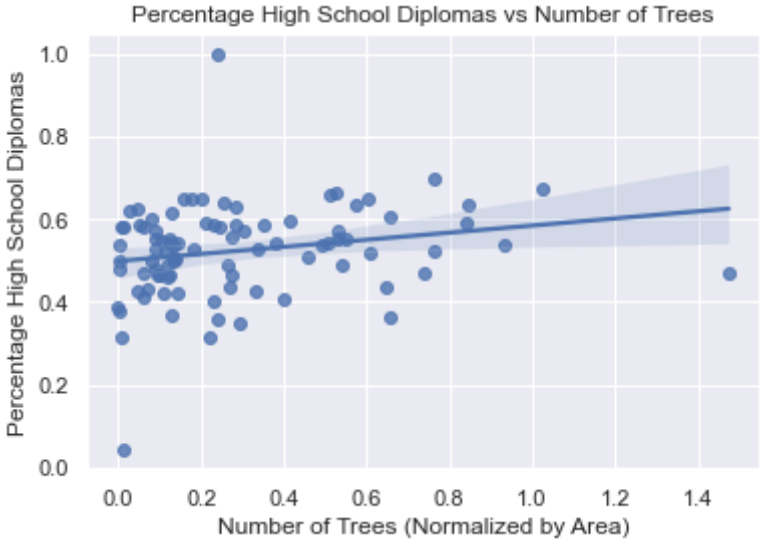


Inference: Interestingly, we do observe a positive correlation between the percentage of commercial areas in a neighborhood and the tree density.

Education and Trees

In [85]:
```
plot = sns.regplot(x = 'area_norm_tree_count', y = 'per_diploma', data = complete_data)
plot.set(xlabel = "Number of Trees (Normalized by Area)", ylabel = "Percentage High School Diplomas",
         title = "Percentage High School Diplomas vs Number of Trees")
```

Out[85]:
```
[Text(0.5, 0, 'Number of Trees (Normalized by Area)'),
 Text(0, 0.5, 'Percentage High School Diplomas'),
 Text(0.5, 1.0, 'Percentage High School Diplomas vs Number of Trees')]
```



Inference: We observe no correlation between high school education and tree density.

In [ ]: