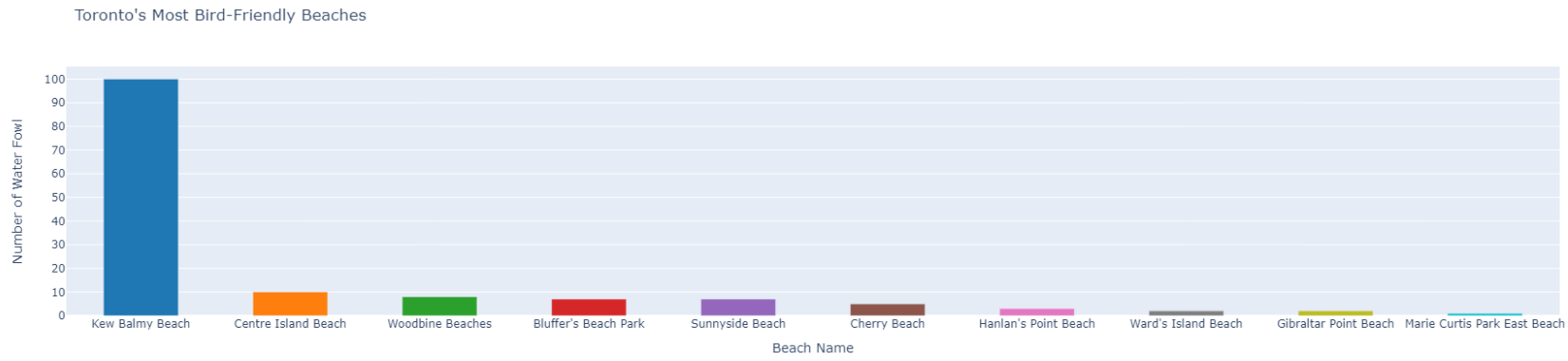
Data Visualization – Assignment 4

Visualization 1:

Toronto Beaches Observations – https://open.toronto.ca/dataset/toronto-beaches-observations/



What software did you use to create your data visualization?

- Python.

Who is your intended audience?

- Bird lovers in Toronto.

What information or message are you trying to convey with your visualization?

- Beaches to visit for waterfowl enthusiasts.

What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?

- From an aesthetic perspective, I used a different colour for each beach. More importantly, the graph clearly shows the main beach to visit (Kew Balmy Beach) if one is so inclined to watch waterfowl.

How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization?

- I used publicly available data and provided the Python code used to generate the visualization.

How did you ensure that your data visualization is accessible?

- Each destination is colour-coded, but text labels are included for those with visual impairments (e.g. colour blindness).

Who are the individuals and communities who might be impacted by your visualization?

- Mostly Toronto communities (where the various beaches are located).

How did you choose which features of your chosen dataset to include or exclude from your visualization?

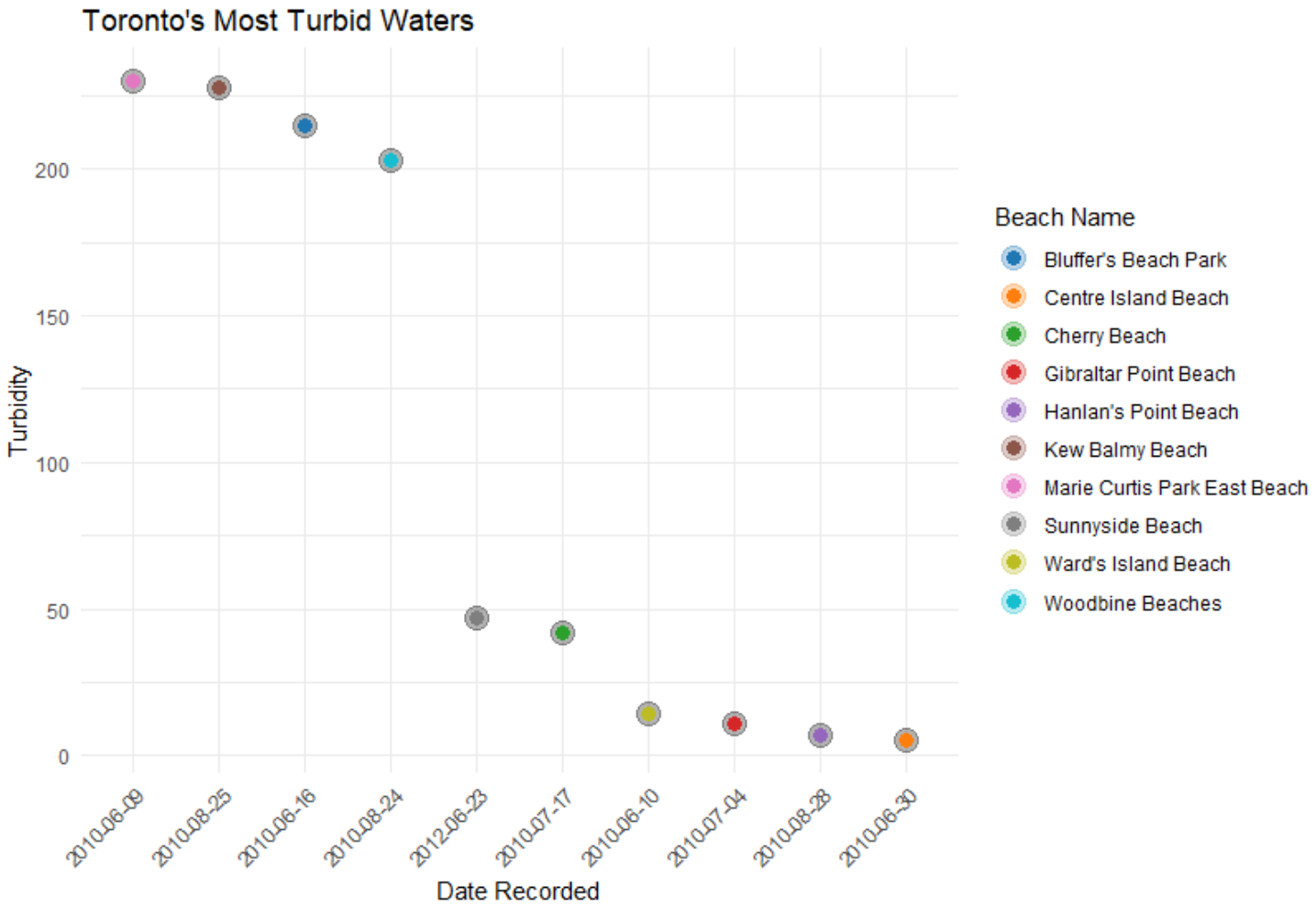
- I wanted a chart that would be valuable to a specific community, so I singled out an obscure but helpful metric, in this case, the number of waterfowl observed at each of Toronto's beach locations.

What 'underwater labour' contributed to your final data visualization product?

- The efforts of whoever contributed to Toronto's Open Data project.
- Contributors to Python and its various packages.
- My Python, and Data Viz teachers.

Visualization 2:

Toronto Beaches Observations – https://open.toronto.ca/dataset/toronto-beaches-observations/



What software did you use to create your data visualization?

- R.

Who is your intended audience?

- Toronto beach goers.

What information or message are you trying to convey with your visualization?

- Beaches that have recorded the worst water turbidity.

What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?

- I used both substantive and perceptual approaches to highlight the unusually high turbidity numbers recorded at 4 beaches in 2010.

How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization?

- I used publicly available data and provided the R code used to generate the visualization.

How did you ensure that your data visualization is accessible?

- I used accessible colours from the Viridis Color Palette Generator.

Who are the individuals and communities who might be impacted by your visualization?

- Toronto beach goers, tourists, and Toronto Parks, Forestry & Recreation staff members.

How did you choose which features of your chosen dataset to include or exclude from your visualization?

- I chose turbidity and location to show how bad turbidity was and where. I also chose to include dates to show that these incidents took place at a certain point in time and are not necessarily ongoing.

What ‘underwater labour’ contributed to your final data visualization product?

- The efforts of whoever contributed to Toronto’s Open Data project.
- Folks from Toronto’s Parks, Forestry & Recreation division.
- Contributors to R, including the people who contributed to the various packages I used to plot my data.
- My R, and Data Viz teachers!

Appendix A: Python code

```
#Import and install all necessary packages
!pip install scipy
!pip install PIL
!pip install requests
!pip install plottable
!pip install plotly



import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import scipy
import PIL
import requests
from plottable import Table, ColumnDefinition
```

```
import textwrap as twp
import plotly.graph_objects as go

#Import CSV file
df = pd.read_csv('C:/Users/busyf/Downloads/toronto-beaches-observations.csv', on_bad_lines='skip')

#View and inspect data
df



# Sort data by date, then by number of water fowl, and finally, remove duplicates and rows with empty values
df_sorted = df.sort_values(by=["dataCollectionDate", "waterFowl"],
ascending=False).dropna(subset=["waterFowl"]).drop_duplicates(subset=["beachName"])

# Select top 10 results
df_top10 = df_sorted.head(10)

# Sort the top 10 results by the "waterFowl" column
df_top10 = df_top10.sort_values(by="waterFowl", ascending=False)
```

```
# Plot the top 10 beaches with the highest water fowl populations
x1 = df_top10["beachName"]
y1 = df_top10["waterFowl"]



#Different colours for the each beach
palette = ["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd", "#8c564b", "#e377c2", "#7f7f7f", "#bcbd22",
"#17becf"]


graph = go.Figure()
graph.add_trace(go.Bar(x=x1, y=y1,marker_color=palette, width=0.5))
graph.update_layout(
    title="Toronto's Most Bird-Friendly Beaches",
    xaxis_title="Beach Name",
    yaxis_title="Number of Water Fowl"


    )
```

```
# Increase ticks on y-axis
graph.update_yaxes(
    tickmode='linear',
    tick0=0,
    dtick=10
)
```

Appendix B: R code

```
# Install and load packages
library(ggplot2)
library(dplyr)



# Import CSV file
df <- read.csv("C:/Users/busyf/Downloads/toronto-beaches-observations.csv", stringsAsFactors = FALSE)

# Sort data by date, then by air temperature, and finally, remove duplicates and rows with empty values
df_sorted <- df %>%
  dplyr::arrange(desc(turbidity), desc(dataCollectionDate)) %>%
  dplyr::filter(!is.na(turbidity)) %>%
  dplyr::distinct(beachName, .keep_all = TRUE)

# Select top 10 results
df_top10 <- head(df_sorted, 10)
```

```
# Assign Viridis colors
palette <- c("#fde725", "#b5de2b", "#6ece58", "#35b779", "#1f9e89", "#26828e", "#31688e", "#3e4989", "#482878",
"#440154")


# Plot scatter plot with fancy markers
graph <- ggplot(df_top10, aes(x = reorder(dataCollectionDate, -turbidity), y = turbidity, color = beachName)) +
  geom_point(aes(color = NULL), size = 5, alpha = 0.3) +  # Shadow
  geom_point(size = 3) +  # Actual points

  scale_color_manual(values = palette) +
  labs(title = "Toronto's Most Turbid Waters", x = "Date Recorded", y = "Turbidity", color = "Beach Name") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

graph
```