```
---
title: "DANL 310: Data Visualization and Presentation\nMidterm Exam"
date: "2023-04-06"
---


# Honor Pledges

I solemnly swear that I will not cheat or engage in any form of
academic dishonesty during this exam.

<br>

I will not communicate with other students or use unauthorized
materials.

<br>

I will uphold the integrity of this exam and demonstrate my own
knowledge and abilities.

<br>

By taking this pledge, I acknowledge that academic dishonesty
undermines the academic process and is a violation of the trust placed
in me as a student.

<br>

I accept the consequences of any violation of this promise.

<br>

Student's Signature:



## Exam Questions

The midterm exam questions are provided in the following webpage:
- [https://bcdanl.github.io/DANL310-midterm-q.html](https://
bcdanl.github.io/DANL310-midterm-q.html).


## Loading R packages

```{r, warning=F, message=F}
library(knitr)
library(tidyverse)
library(lubridate)
```
```

```
library(socviz)
library(ggthemes)
library(hrbrthemes)
library(gapminder)
```

- `hrbrthemes::theme_ipsum()` or `theme_minimal()` can be used for your ggplot theme.

```{r setup, include = FALSE}
# library(tidyverse)
# library(knitr)
# library(kableExtra)
# library(ggthemes)
# library(ggtech)
# library(ggthemr)
# library(ggthemes)
# library(hrbrthemes)
# library(hexbin)
# library(ggforce)
# library(skimr)

knitr::opts_chunk$set(fig.width=8, fig.height=5,
                      echo = T, eval = T, warning = F, message = F)
```


Question 1

The following data is for Question 1:

```{r}
gapminder <- gapminder
```

```{r, result = 'asis', echo = F}
rmarkdown::paged_table(gapminder)
```


Q1a. Replicate the given ggplot.

  - Use the color `#0072B2` for dots.

```{r, fig.height= 20, fig.width= 15}
cont_gapminder <- gapminder %>%
```

```
  group_by(continent, country)

ggplot(data = gapminder, aes(x = country, y = lifeExp, color =
continent)) +
  geom_point(data = cont_gapminder) +
  facet_grid(~continent) +
  coord_flip() +
  theme_bw()
```

Q1b. Make a simple comment on the visualization result.

African Countries tend to have lowest life expectancy out of all the continents while the Americas and Europe tend to have the highest life expectancies.

Question 2

The following data is for Question 2:

```{r}
n_tweets_long <- read_csv(
  'https://bcdanl.github.io/data/n_tweets_long.csv')
```

```{r, result = 'asis', echo = F}
rmarkdown::paged_table(n_tweets_long)
```


Q2a. Replicate the given ggplot.

  - The following describes the `type` values:
    - `n_ot_us`: Number of US tweets
    - `n_ot_wrld`: Number of worldwide tweets
    - `n_rt_lk_us`: Number of US retweets & likes
    - `n_rt_lk_wrld`: Number of worldwide retweets & likes
  - Use the colors, `maroon` and `#428bca` properly.

```{r, fig.height, fig.width}
n_tweets_long <- n_tweets_long %>%
  mutate(likes = ifelse(
n_ot_us == 0,
n_ot_wrld == 0))

ggplot(data = n_tweets_long, aes(x = year, y = type)) +
  geom_bar() +
  geom_line()
```

```

Q2b. Make a simple comment on the visualization result.

Over time there has been a increases in retweets and likes on Twitter, but there are more of them worldwide than the US.

<br><br>

# Question 3

The following data set is for Question 3:

```{r}
nyc_dog_license <- read_csv(
  'https://bcdanl.github.io/data/nyc_dog_license.csv')
nyc_zips_coord <- read_csv(
  'https://bcdanl.github.io/data/nyc_zips_coord.csv')
nyc_zips_df <- read_csv(
  'https://bcdanl.github.io/data/nyc_zips_df.csv')
```

```{r, result = 'asis', echo = F}
rmarkdown::paged_table(nyc_dog_license)
```

```{r, result = 'asis', echo = F}
rmarkdown::paged_table(nyc_zips_coord)
```

```{r, result = 'asis', echo = F}
rmarkdown::paged_table(nyc_zips_df)
```


Q3a. Replicate the given ggplot.

  - You should calculate the proportion of `Pit Bull (or Mix)` for each zip code.
  - You should join data.frames properly.
  - Choose the color palette from the `viridis` scales https://ggplot2.tidyverse.org/reference/scale_viridis.html.
  - Use `coord_map(projection = "albers", lat0 = 39, lat1 = 45)`.

```{r, fig.width= , fig.height= }
nyc_dog_license_sum <-

```
nyc_pitbull_map <- left_join(nyc_zips_coord, nyc_zips_df)

nyc_full_map <- left_join(nyc_pitbull_map, nyc_dog_license_sum)
 # ^^^ My R keeps crashing trying to run the code merging my data sets
q3 <- ggplot(data = nyc_full_map) +
  geom_polygon(mapping = aes(x = X, y = Y))

q33 <- q3 + scale_fill_gradient()
```

Q3b. Which `zip_code` does have the highest proportion of `Pit Bull (or Mix)`?

The zip code by Queens has the highest percentage of Pit Bulls.

Question 4

The following data is for Question 4:

```{r}
stock <- read_csv('https://bcdanl.github.io/data/stocks2013_2023.csv')
```

```{r, result = 'asis', echo = F, message = F, warning = F}
rmarkdown::paged_table(stock)
```


Q4a. Replicate the given ggplot.

```{r, fig.height=  10, fig.width= 45}
stock <- stock %>%
  group_by(company, Date)

q4 <- ggplot(stock, aes(x = Date, y = (log(Close)), color = company))
+ geom_point() +
  facet_grid(~company) +
  geom_line(mapping = aes(x = Date, y = (log(Close))))

q4

```

Q4b. In 2020, which company's stock trading `Volume` does seem to be

the most insensitive to a change in `Close` price?

I would say Microsoft is the most insensitive out of all the companies.