CODE

Data Visualization Project 1 Infographic

Examining Road Traffic Speed and Crash Data in Massachusetts from 2019 to 2020

Emmett Greenberg
Northeastern University
INSH 5302: Information Design and Visual Analytics

January 30, 2025

Introduction

> How did traffic patterns in Massachusetts change from 2019 to 2020, especially during the onset of the COVID-19 pandemic?

I analyzed changes in traffic occurring on roads in Massachusetts in 2020 versus 2019. I created visualizations showing changes in speed and the number of car accidents by road type (interstate, arterial, local road, etc.), using boxplots to show the distributions of each of these two variables. I also wanted to know which communities in

Massachusetts saw the biggest increase in accidents in 2020, so I created a bar chart of the top 10 results, grouped by the name of the city or town the road is in. I wanted to display trends occurring on different types of roads and to look for correlations between speed and accidents.

Data

Load Libraries and Packages

```
library(mosaic)
library(tidyverse)
library(moderndive)
library(plotly)
```

Description of data

> R Package: moderndive

> Data set: I chose the ma_traffic_2020_vs_2019 dataset found in the moderndive package. This dataset contains information about changes in speed, volume, and accidents of car traffic between 2020 and 2019 by community and class of road in Massachusetts.I chose this data set because of my interest in studying transportation and mobility. I am interested in learning how traffic patterns change over time, and the onset of COVID-19 pandemic in 2020 had a major impact on transportation choices throughout Massachusetts.

Load and Clean Data

```
# load data
ma_traffic_2020_vs_2019 <- moderndive::ma_traffic_2020_vs_2019
attach(ma_traffic_2020_vs_2019)

# clean data
# check for NAs
colSums(is.na(ma_traffic_2020_vs_2019))
```

```
##           community    functional_class     change_in_speed    change_in_volume
##                   0                   0                   0                   2
## change_in_accidents
##                  78
```

```
# remove all nulls (NA)
ma_traffic_2020_vs_2019 <- ma_traffic_2020_vs_2019 %>% na.omit()

# check for NAs again
colSums(is.na(ma_traffic_2020_vs_2019))
```

```
##           community    functional_class     change_in_speed    change_in_volume
##                   0                   0                   0                   0
## change_in_accidents
##                   0
```

Preview Data

```
head(ma_traffic_2020_vs_2019, n=10)
```

community	functional_class	change_in_speed	change_in_volume	change_in_accidents
abington	Arterial	3.7112376	-72554	-60
acushnet	Arterial	-2.3260001	7477	-11
acushnet	Collector	0.8824872	9515	2
adams	Arterial	3.9161216	-292144	-4
agawam	Arterial	1.9159067	-49674	-8
agawam	Freeway & Expressway	8.2742166	4867679	-12
amesbury	Freeway & Expressway	-13.6267960	13530	-3
amherst	Arterial	2.4639441	5369	5
andover	Arterial	-10.4378544	-2239795	-85
andover	Interstate	28.1457775	85770620	-5

Variables

```
names(ma_traffic_2020_vs_2019)
```

```
## [1] "community"           "functional_class"    "change_in_speed"
## [4] "change_in_volume"    "change_in_accidents"
```

The variables featured in this data visualization project are:

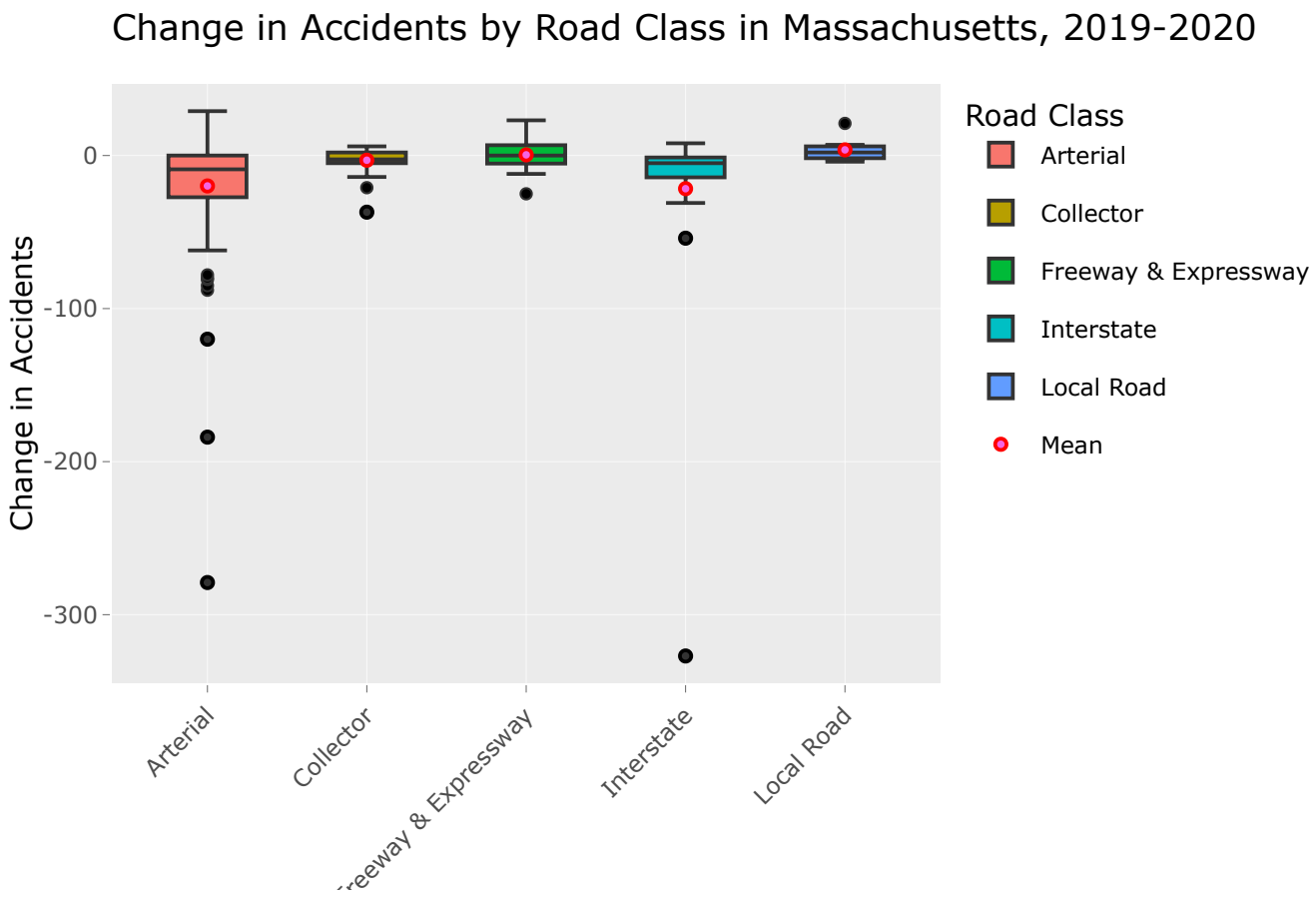
1. **community**: name of the city or town
2. **functional_class**: class or group the road belongs to
3. **change_in_speed**: change in average estimated speed (mph) from 2019 to 2020
4. **change_in_volume**: change in traffic volume (number of vehicles recorded at each site) from 2019 to 2020
5. **change_in_accidents**: change in number of accidents at each site from 2019 to 2020

Data Analysis

Required Analysis

Chart/Graph

```
# Boxplot of change in accidents by road type
p <- ggplot(ma_traffic_2020_vs_2019, aes(x = functional_class, y = change_in_accidents, fi
ll = functional_class)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Change in Accidents by Road Class in Massachusetts, 2019-2020",
       x = "Road Class",
       y = "Change in Accidents",
       fill = "Road Class")

p <- p + stat_summary(fun=mean, geom = "point", color = "red", aes(fill="Mean"))


# make it interactive
interactive_p <- ggplotly(p)

# display
interactive_p
```



Road Class

Summary Statistics

```
favstats(~change_in_accidents, data=ma_traffic_2020_vs_2019)
```

min	Q1	median	Q3	max	mean	sd	n	missing
-327	-15.25	-5	1	29	-14.86957	39.35052	184	0

```
favstats(change_in_accidents ~ functional_class, data=ma_traffic_2020_vs_2019)
```

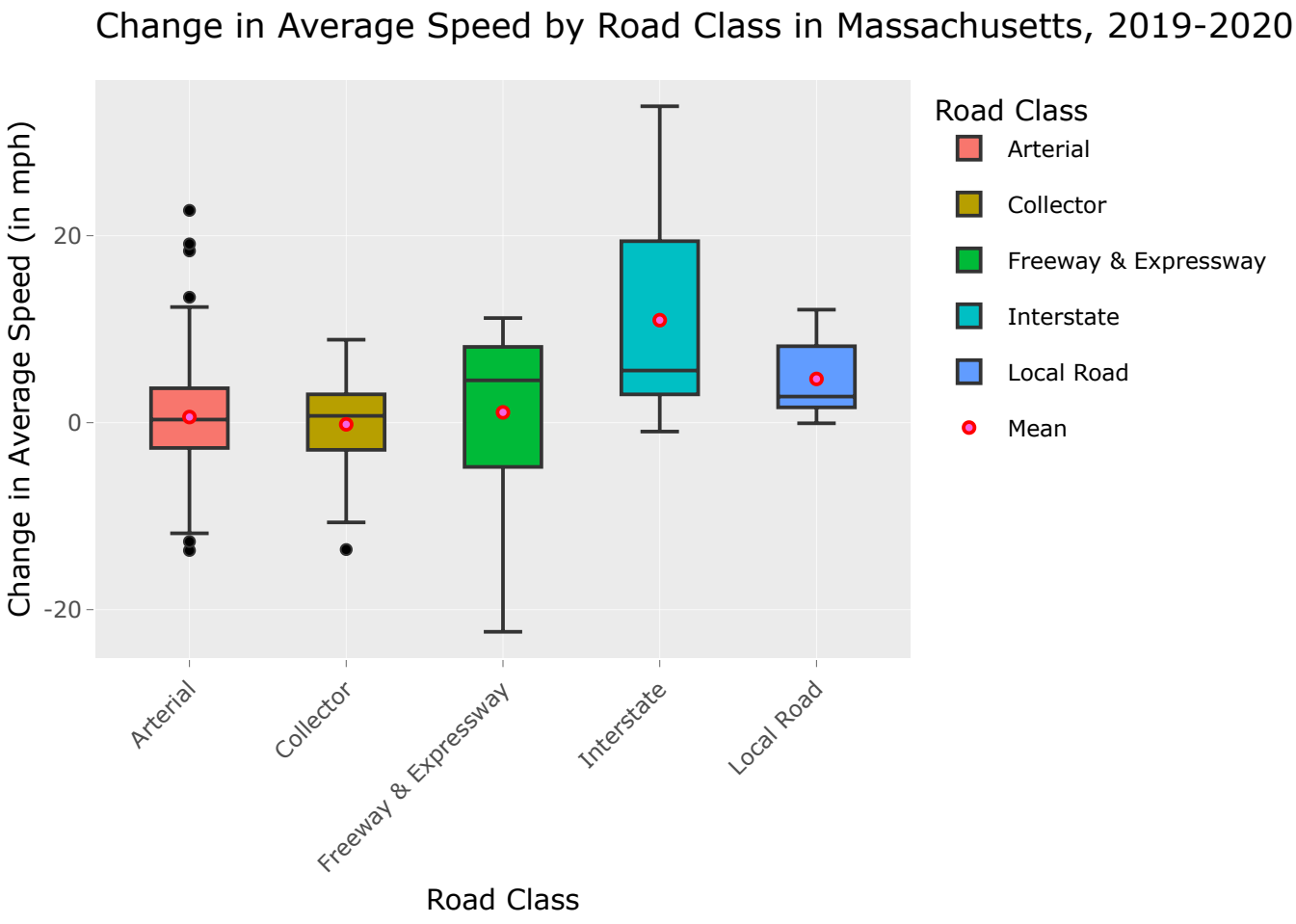
functional_class	min	Q1	median	Q3	max	mean	sd	n	missing
Arterial	-279	-27.0	-9.0	0.0	29	-19.8623853	38.976362	109	0
Collector	-37	-5.0	-2.5	2.0	6	-3.0588235	8.416942	34	0
Freeway & Expressway	-25	-4.5	0.0	5.5	23	0.4545455	12.941125	11	0
Interstate	-327	-13.5	-5.0	-1.5	8	-21.6521739	67.870605	23	0
Local Road	-4	-1.5	2.0	5.0	21	3.7142857	8.440266	7	0

Additional Analyses

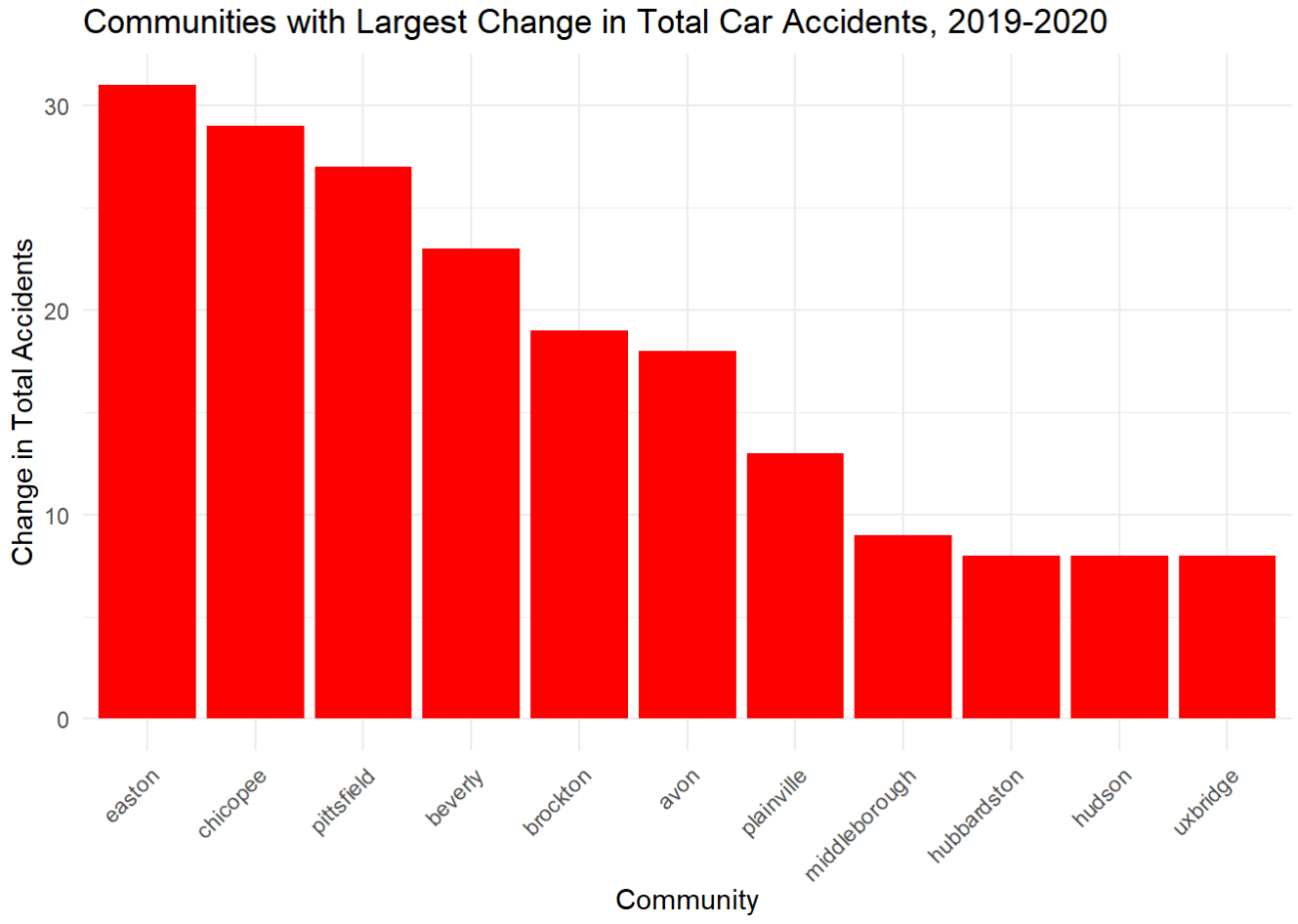
```
favstats(change_in_speed ~ functional_class, data = ma_traffic_2020_vs_2019)
```

functional_class	min	Q1	median	Q3	max	mean	sd	n	missing
Arterial	-13.6690379	-2.701585	0.3331846	3.671656	22.710283	0.6024685	6.419604	109	0
Collector	-13.5736297	-2.779989	0.7288866	2.807465	8.877342	-0.1781318	4.640126	34	0
Freeway & Expressway	-22.3708736	-2.092891	4.5272962	7.936091	11.190034	1.1044626	10.827723	11	0
Interstate	-0.9514444	3.028124	5.5771465	19.158442	33.850074	10.9663826	11.405843	23	0
Local Road	-0.0646388	2.012336	2.7932468	6.955665	12.089780	4.6791986	4.446985	7	0

```
# Boxplot of change in speed by road type
q <- ggplot(ma_traffic_2020_vs_2019, aes(x = functional_class, y = change_in_speed, fill =
functional_class)) +
  geom_boxplot(show.legend = FALSE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Change in Average Speed by Road Class in Massachusetts, 2019-2020",
       x = "Road Class",
       y = "Change in Average Speed (in mph)",
       fill = "Road Class"
       )

# add points for mean
q <- q + stat_summary(fun=mean, geom = "point", color = "red", aes(fill="Mean"))

# make it interactive
interactive_q <- ggplotly(q)

interactive_q
```

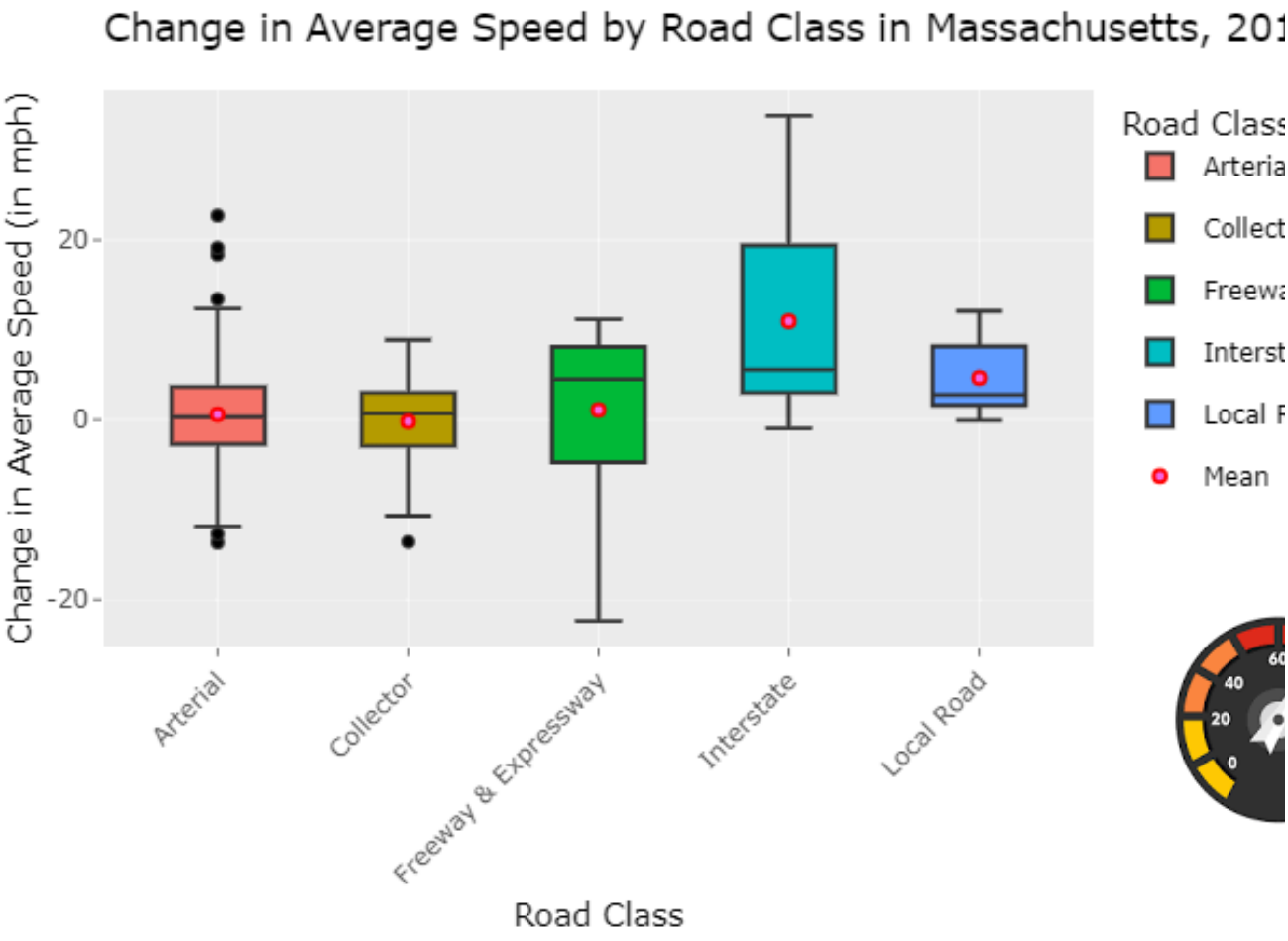


```
# communities with the most accidents
most_accidents <- ma_traffic_2020_vs_2019 %>%
  group_by(community) %>%
  summarise(total_sites=n(),
            total_change_in_accidents=sum(change_in_accidents)) %>%
  slice_max(total_change_in_accidents, n=10)

# bar chart of communities ordered
most_accidents_plot <- ggplot(most_accidents, aes(x=reorder(community, -total_change_in_ac
cidents), y=total_change_in_accidents)) +
  geom_bar(stat = "identity", fill="red") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Communities with Largest Change in Total Car Accidents, 2019-2020",
       x = "Community",
       y = "Change in Total Accidents")

most_accidents_plot
```



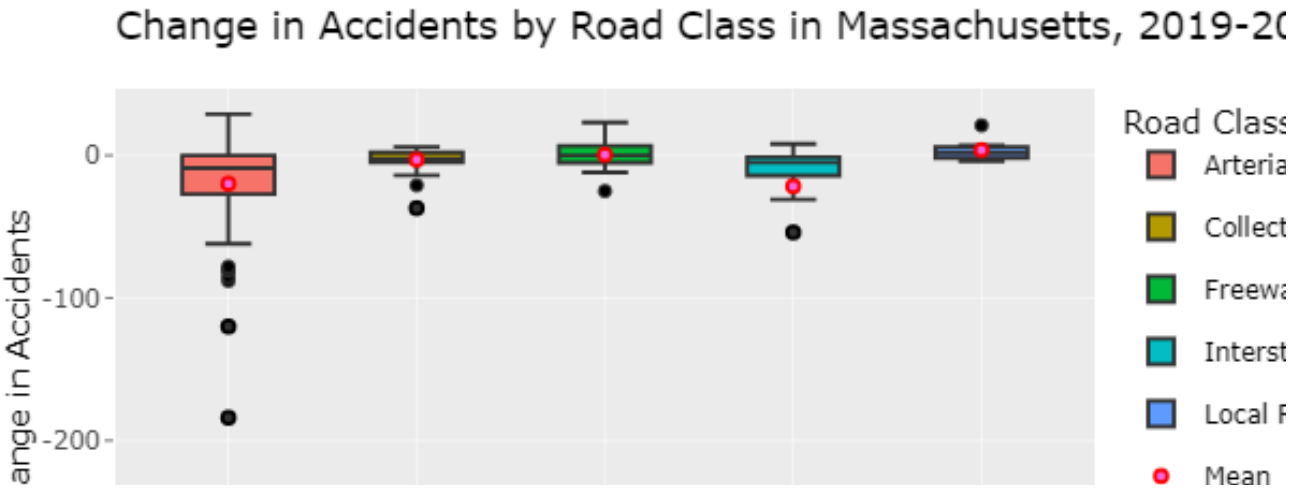
Infographic

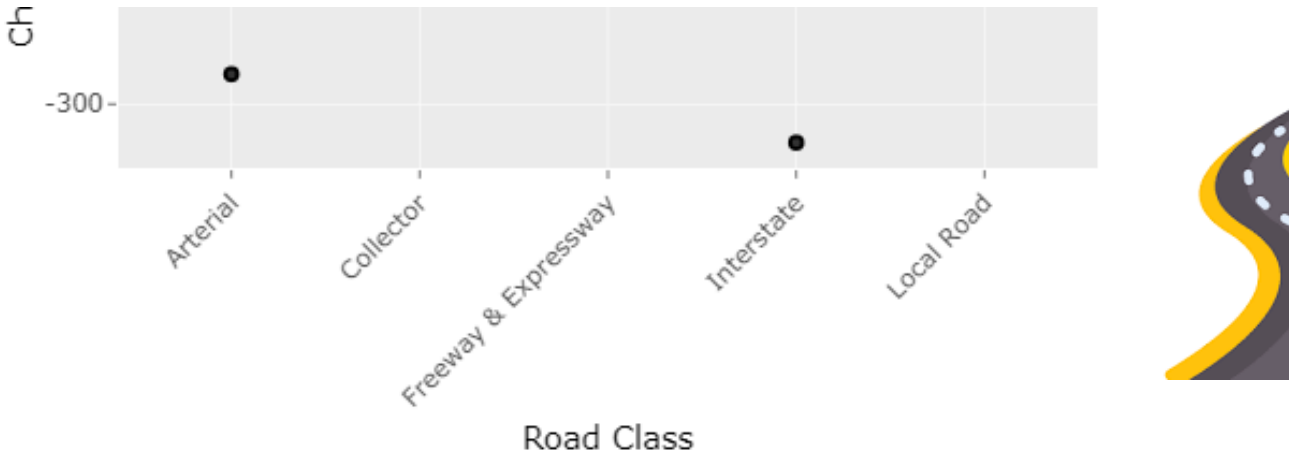
communities



Most roads saw a **small to moderate increase** in average car speed in 2020.

Interstates had, on a 10.97 mph incre average car spe



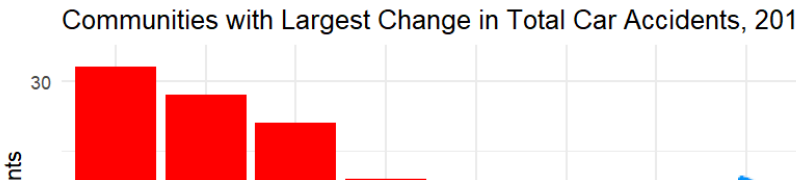


On average, most road types saw **fewer accidents** in 2020.

However, **local** averaged **about tw accidents** in 20 addition to averag speeds of 4.68

Community Spotlight

The three



Midterm Infographic (https://www.canva.com/design/DAGMLI0yeOA/jEqtyisHM034pwFUrfjGwA/view?utm_content=DAGMLI0yeOA&utm_campaign=designshare&utm_medium=embeds&utm_source=link) by Emmett Greenberg

URL Link to Canva Site: Emmett Greenberg - Data Viz Project 1 (https://www.canva.com/design/DAGMLI0yeOA/jEqtyisHM034pwFUrfjGwA/view)

References

1. https://cran.r-project.org/web/packages/moderndive/moderndive.pdf (https://cran.r-project.org/web/packages/moderndive/moderndive.pdf)
2. https://www.statology.org/boxplot-with-mean-in-r/ (https://www.statology.org/boxplot-with-mean-in-r/)

3. https://plotly.com/ggplot2/box-plots/ (https://plotly.com/ggplot2/box-plots/)