MINORITY DEMOGRAPHICS (BLACK, ASIAN, HISPANIC/LATIN)

Population Information for Top 40 Polulated Cities and Corresponding Regions
Source: US 2020 Census
Jupyter Notebooks: Google Drive

- HC_Poli_EDA.ipynb

Findings:

For total population estimate number for the Top 40 Cities:

- West has the most people, followed by Northeast, Southwest, Midwest, and then Southeast
- New York has a significantly larger population, followed by Los Angeles, Chicago, Houston

For minority populations in the Top 40 Cities:

- Total Sum of Black, Asian, and Hisp/Lat
    - The Northeast has the largest sum closely followed by the West then Southwest, Midwest, and finally Southeast
- Total Minority Percentages
    - The West closely followed by the Southwest have the largest total percentage of minorities
    - Detroit and El Paso had the highest totals for total minority percentages
- Black Percentages
    - The Northeast has a larger percentage of Black people than any other race/ethnicity
    - West had the lowest percentage of Black people
    - Highest percentages of Black people: **Detroit*** -> Memphis* -> Baltimore* -> Atlanta* -> Washington* -> Philadelphia* -> Milwaukee* -> Charlotte -> Jacksonville -> Chicago
      (asterisk indicates where they are the highest minority percentage compared to the other two)
    - Lowest percentage of Black people: San Jose* -> Albuquerque -> El Paso -> Mesa -> San Francisco ->Tucson -> Portland -> San Diego -> Colorado Springs -> San Antonio

(asterisk indicates where they are the lowest minority percentage compared to the other two)

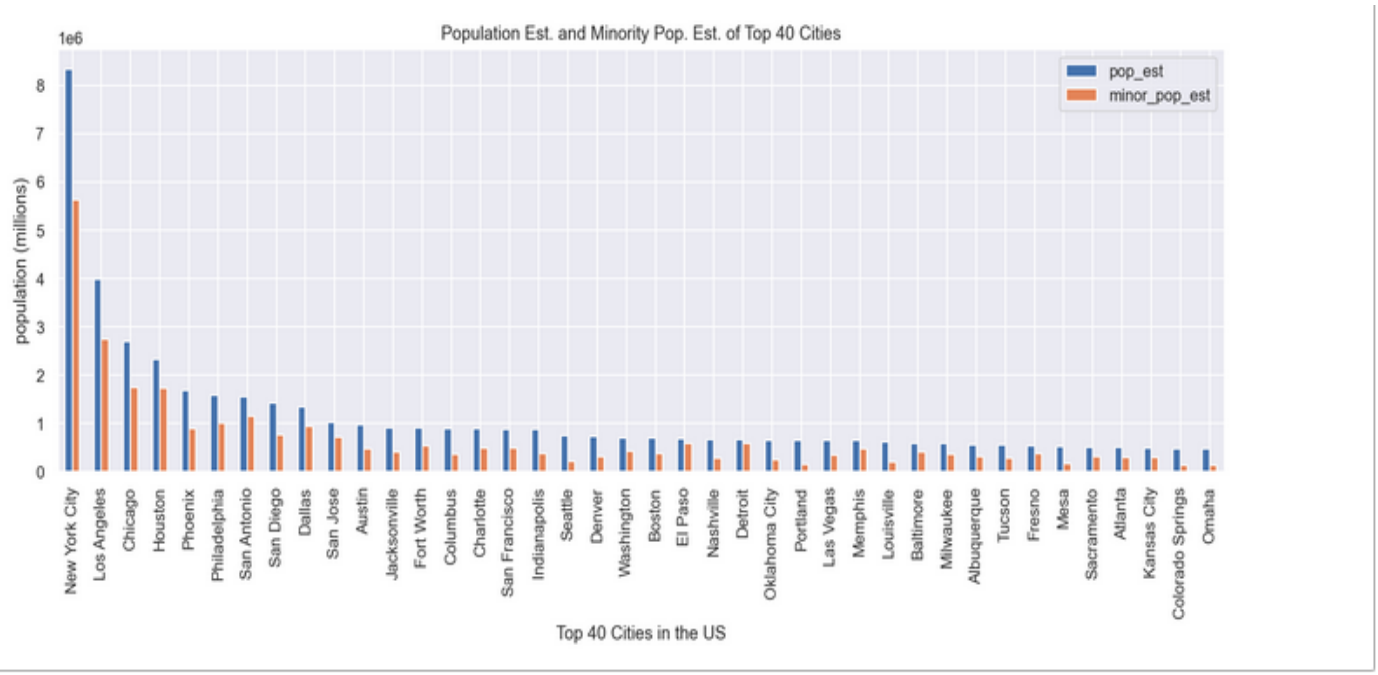
- Hisp/Lat Percentages
    - The Southwest had a larger percentage of Hisp/Lat people than any other race/ethnicity
    - Hispanics were the lowest percentage of minorities in the Southeast
    - Highest percentages of Hisp/Lat people: **El Paso*** -> San Antonio* -> Fresno* -> Albuquerque* -> Los Angeles* -> Houston* -> Tucson -> Phoenix -> Dallas* -> Fort Worth
      (asterisk indicates where they are the highest minority percentage compared to the other two)
    - Lowest percentage of Hisp/Lat people: Atlanta -> Baltimore -> Louisville -> Columbus -> Seattle -> Memphis -> Detroit ->Portland -> Jacksonville -> Nashville
      (asterisk indicates where they are the lowest minority percentage compared to the other two)
- Asian Percentages
    - Lowest percentage in all regions than any other minority, except in the West where Black is the lowest percentage
    - Highest percentages of Asian people: **San Jose*,** San Francisco, Sacramento, San Diego, Seattle, New York City, Fresno, Los Angeles, Boston, Portland
      (asterisk indicates where they are the highest minority percentage compared to the other two)
    - Lowest percentage of Asian people: El Paso* -> Memphis* -> Detroit* -> Mesa* -> Baltimore* -> Louisville* -> San Antonio* -> Albuquerque* -> Colorado Springs* -> Tucson*
      (asterisk indicates where they are the lowest minority percentage compared to the other two)

EDA Method:

Took Census data that contained population estimate and demographic percentages of race/ethnicity

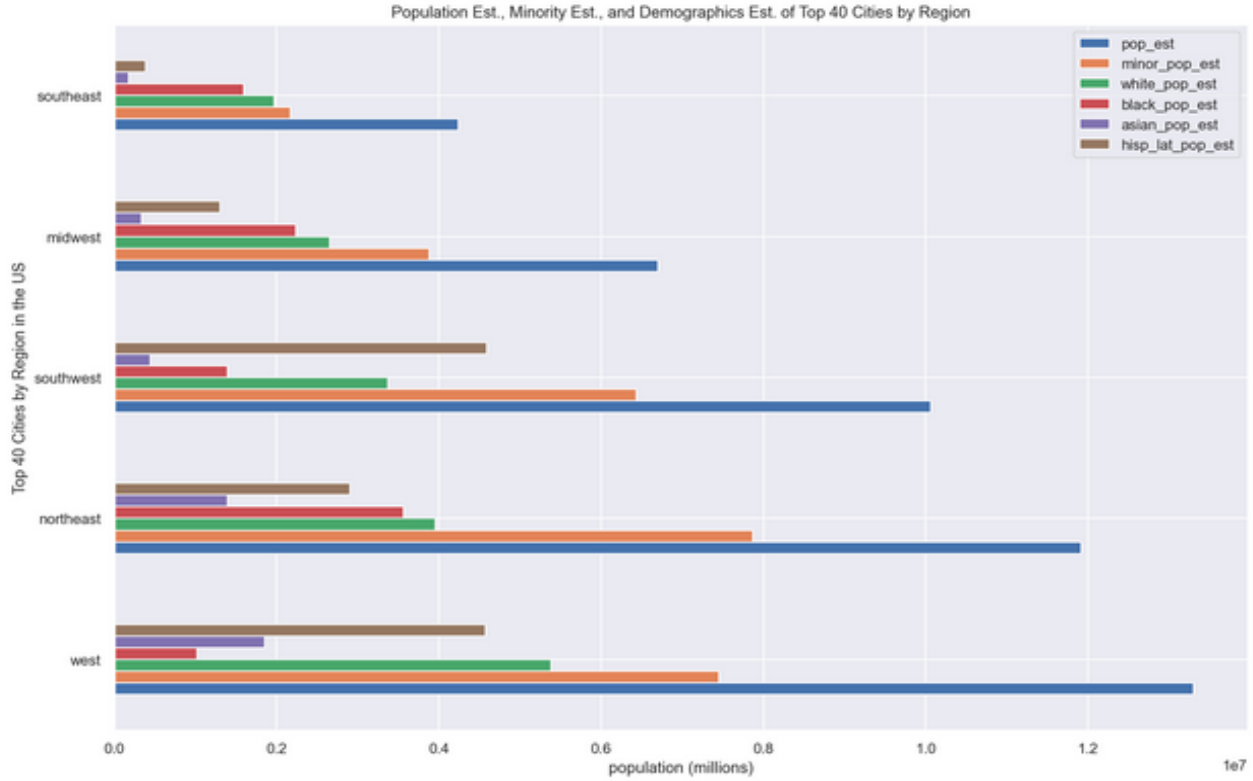
- Filtered for Top 40
- Created features for total minority estimated population, total minority percentage (ratio), and each demographic's population estimate and percentage of the total population for the top 40 cities
- Did various groupby functions with city, region, and the different representations of population and population ratios with functions like .sum() and .sort_values to get the top 10.

Population Estimates (Total and Total Minority) for Top 40 Cities and Region

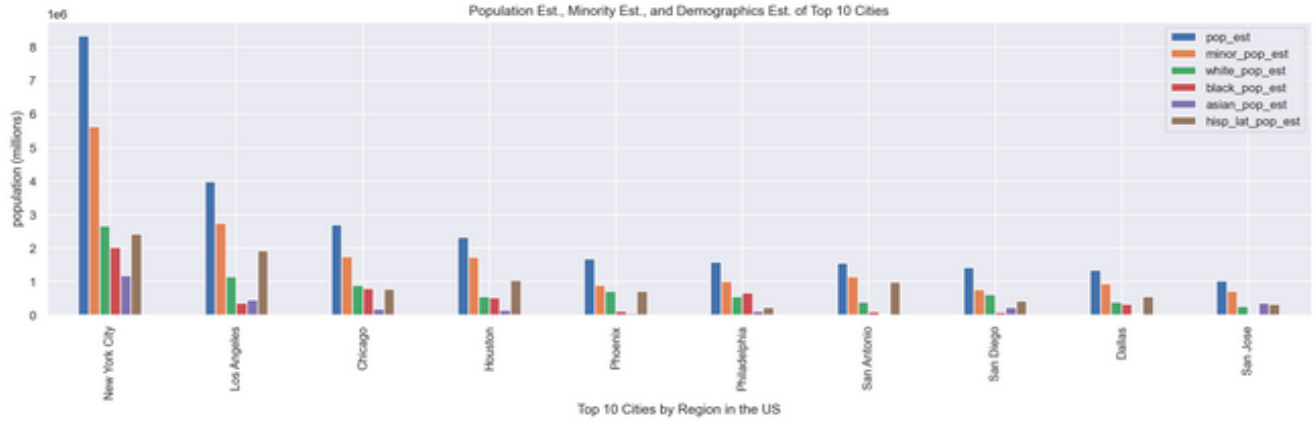


In [165]:
```
pop_region_plot = top_df.groupby("region")[["pop_est", "minor_pop_est"]].sum().sort_values(by=['pop_est'], ascending
pop_region_plot.plot.bar(figsize = (15,5))
plt.xlabel("Top 40 Cities by Region in the US")
plt.ylabel("population (millions)")
plt.title("Population Est. and Minority Pop. Est. of Top 40 Cities by Region");
```

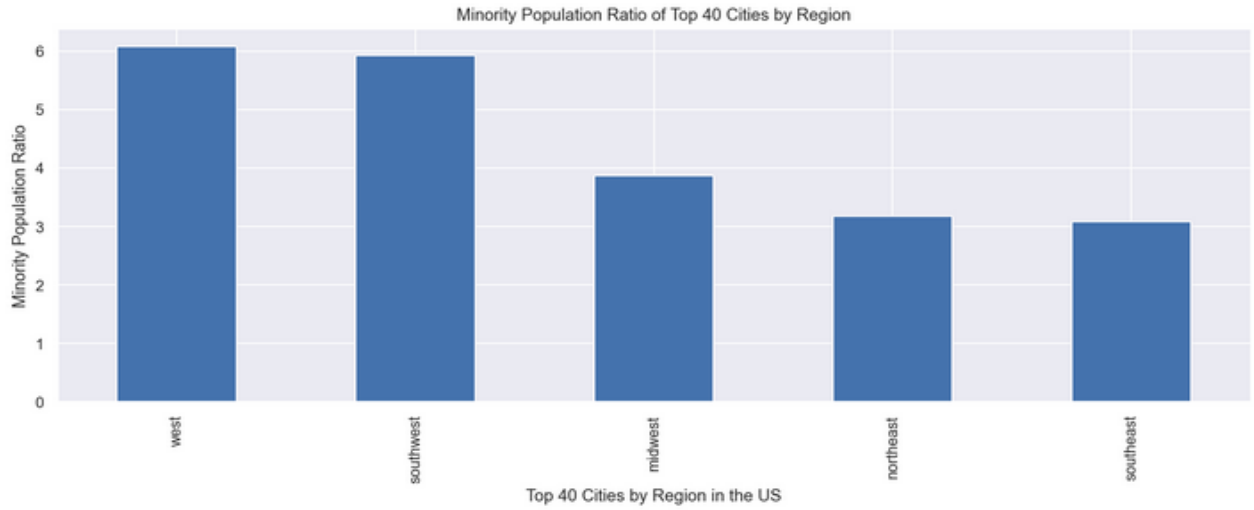
Population Estimates of Total Pop, Total Minorities, White, Black, Asian, and Hisp/Lat by Region and Top 10 Cities (for graph visibility)



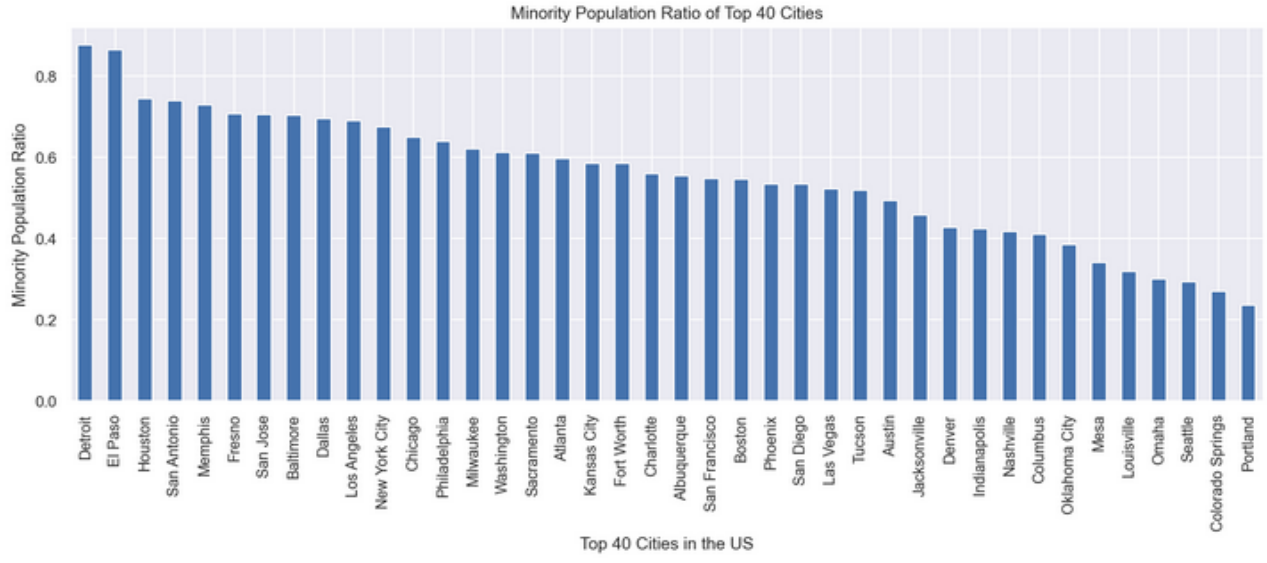

```
[169]: pop_demo_plot = top_df.groupby("city")[["pop_est", "minor_pop_est", "white_pop_est","black_pop_est","asian_pop_est",
pop_demo_plot.plot.bar(figsize = (20,5))
plt.xlabel("Top 10 Cities by Region in the US")
plt.ylabel("population (millions)")
plt.title("Population Est., Minority Est., and Demographics Est. of Top 10 Cities");
```



Total Minority Out of Total Population Percentage for Region and Top 40 Cities

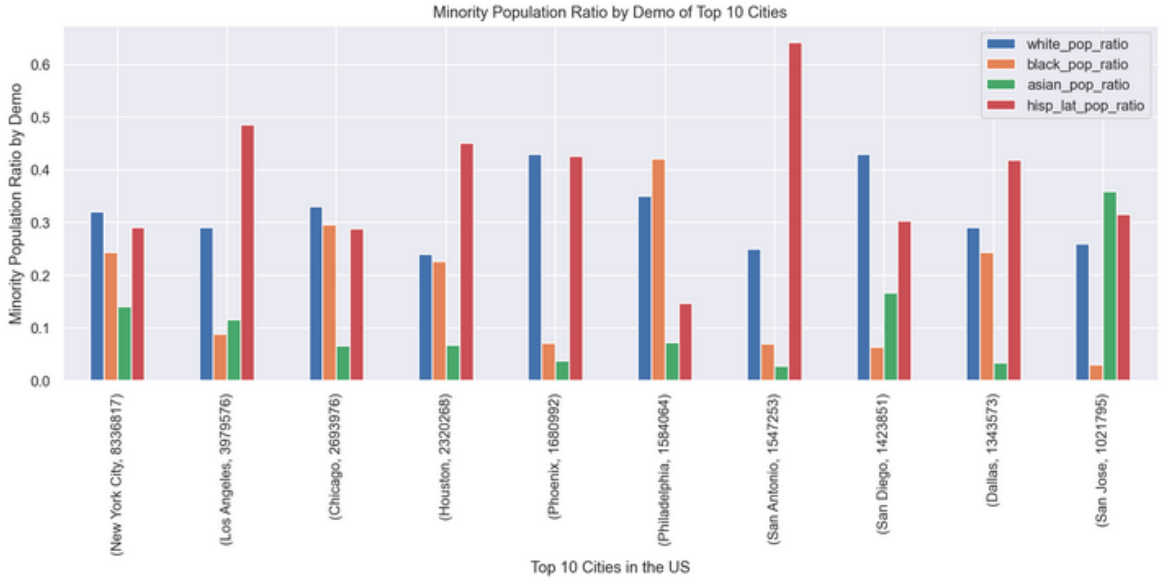



```
]: ratio_city_plot = top_df.groupby("city")["minor_pop_ratio"].sum().sort_values(ascending = False)
   ratio_city_plot.plot.bar(figsize = (15,5))
   plt.xlabel("Top 40 Cities in the US")
   plt.ylabel("Minority Population Ratio")
   plt.title("Minority Population Ratio of Top 40 Cities");
```



Population Percentages Broken Down by Demographic for Region and Top 10 Cities (for graph visibility)

[159]:
```
dem_ratio_city_plot = top_df.groupby(["city","pop_est"])[["white_pop_ratio", "black_pop_ratio", "asian_pop_ratio", "
dem_ratio_city_plot.plot.bar(figsize = (15,5))
plt.xlabel("Top 10 Cities in the US")
plt.ylabel("Minority Population Ratio by Demo")
plt.title("Minority Population Ratio by Demo of Top 10 Cities");
```



HATE CRIMES

Source: [2019 FBI Hate Crime Database]
Jupyter Notebooks: [Google Drive]
- HC_Poli_EDA.ipynb
- Crime_Cleaning.ipynb

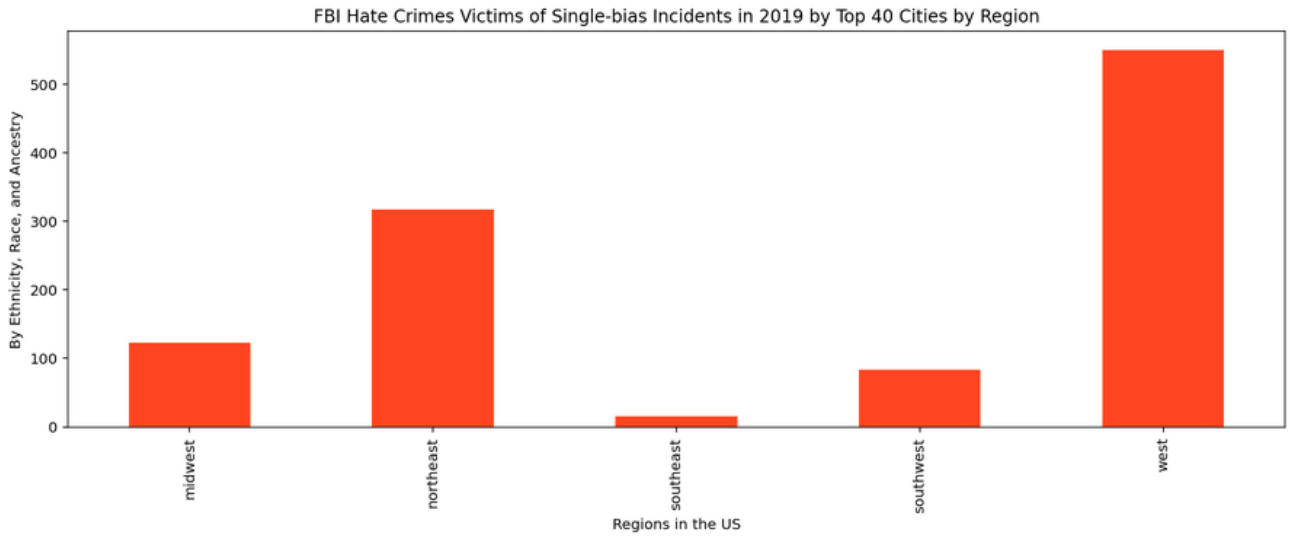
Method:

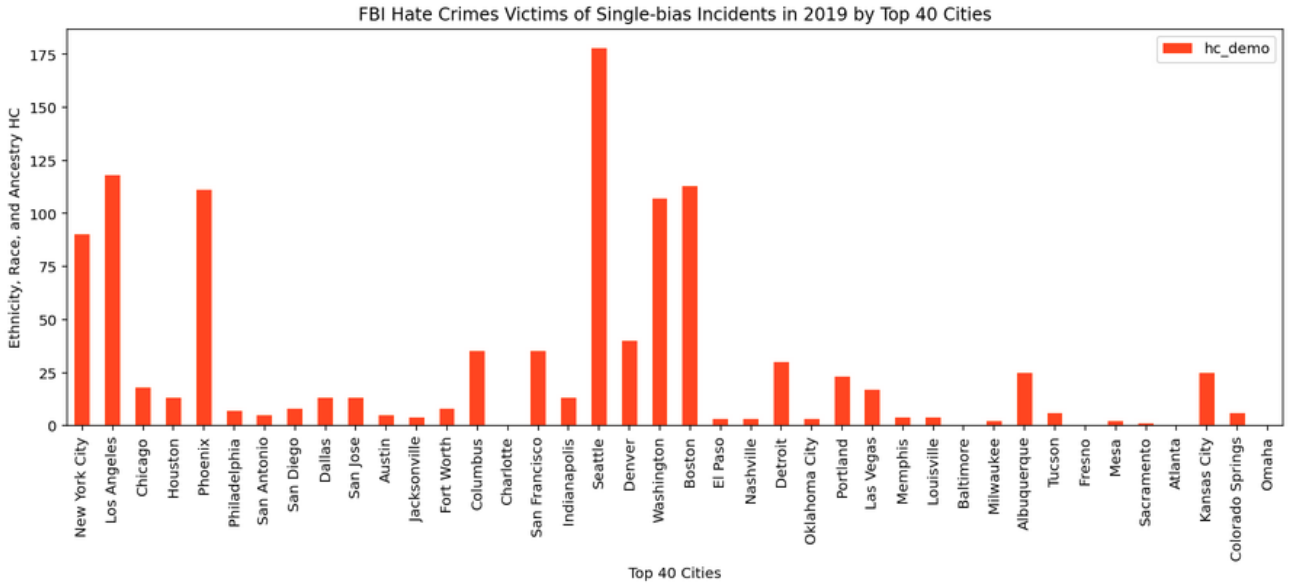
- I took the cleaned database that I had for hate crimes by top 40 city and did a couple groupbys for the bar charts (groupby city and groupby region) along with the sum of hate crimes.

- For the correlation I took the ratio of minority population / total population for each city and ran a correlation function on that number and the number of hate crimes, per city.

Findings:

A large majority of victims from hate crimes in 2019 occured in the West part of the country:
- Seattle being much larger than any other city (this could be due to a single event, that had a lot of victims)
- Seattle -> Los Angeles -> Boston -> Phoenix -> Washington D.C. -> NYC -> Denver -> Columbus -> San Francisco -> Detroit



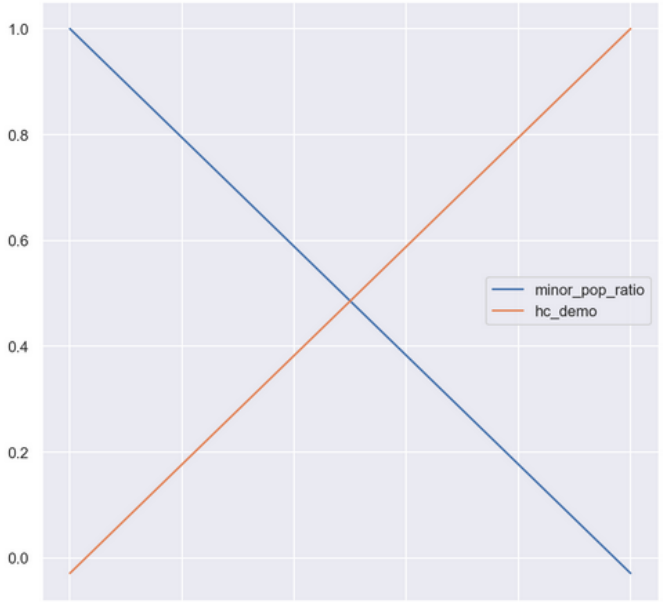


There is a negative correlation (~-0.028896) for percentage of total minorities to hate crime victim counts meaning larger percentages of total minorities in a city correlate with less hate crime victims.

Out[111]:

	minor_pop_ratio	hc_demo
minor_pop_ratio	1.000000	-0.028896
hc_demo	-0.028896	1.000000

```
In [113]: minor_hc_corr.plot.line();
```



POLITICAL PARTIES
(MAYOR, GOVERNOR, STATE)

Source: 2018 House of Representatives | 2020 Senate | 2021 Mayors | 2021 Governors
Jupyter Notebooks: Google Drive

- HC_Poli_EDA.ipynb
- Political_Cleaning.ipynb

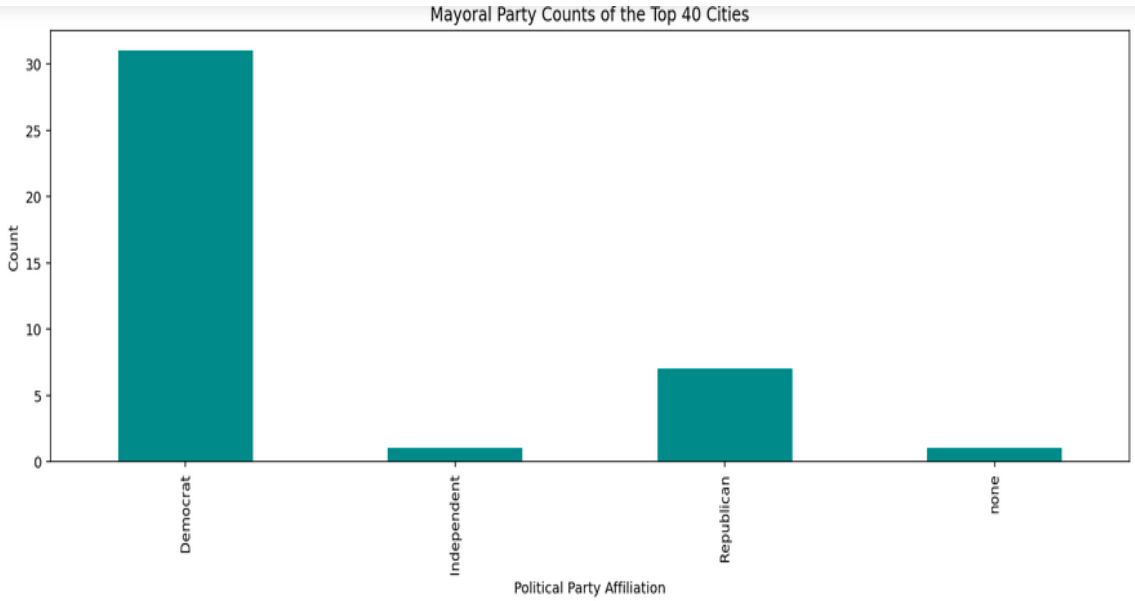
Findings:

- Out of the top 40 cities:
    - Mayors are mostly Democratic
    - Governors are mostly Republican but close to even with Democrat
    - and Congress is mostly Democrat
- Regionally for the top 40 cities:
    - Northeast has all Democrat mayors, Southwest has most of the Republican mayors
    - West has a large majority of Democrat governors, Southwest has a large majority of Republican governors, Northeast has an even amount of Democrat and Republican governors
    - West and Northeast have all Democrat majority states in Congress; Southwest, Southeast, and Midwest have all Republican majority states in Congress

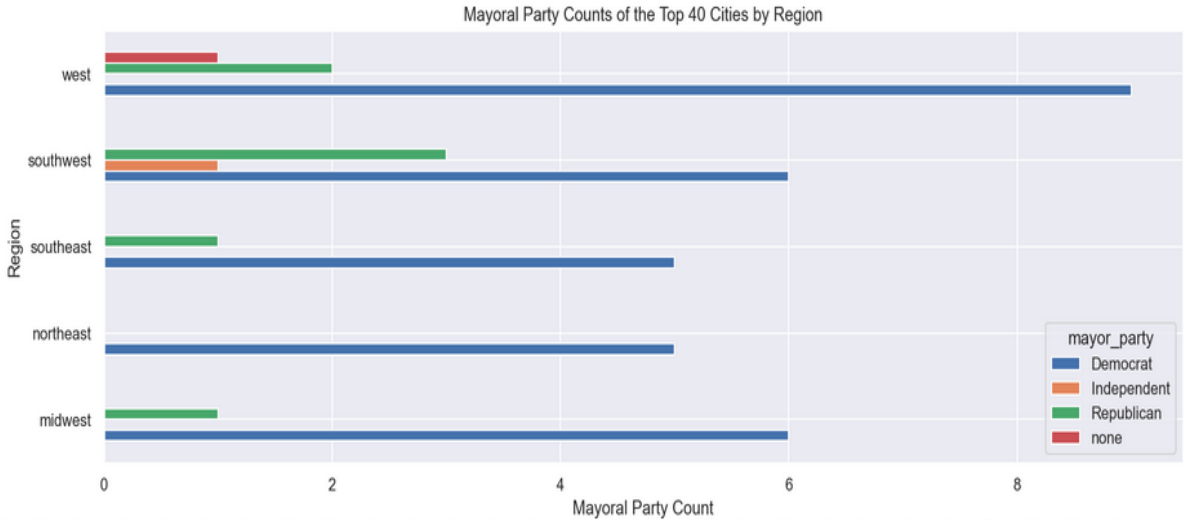
Method:

- I took a cleaned database with the top 40 cities which had features categorizing the political party for the mayor, governor, and congress majority for each city.
- For the count graphs: I took the city and grouped by the political party count for each type (mayor, governor, congress majority)
- For the regional graphs: I took the value_counts of the political party and grouped by the region for each type

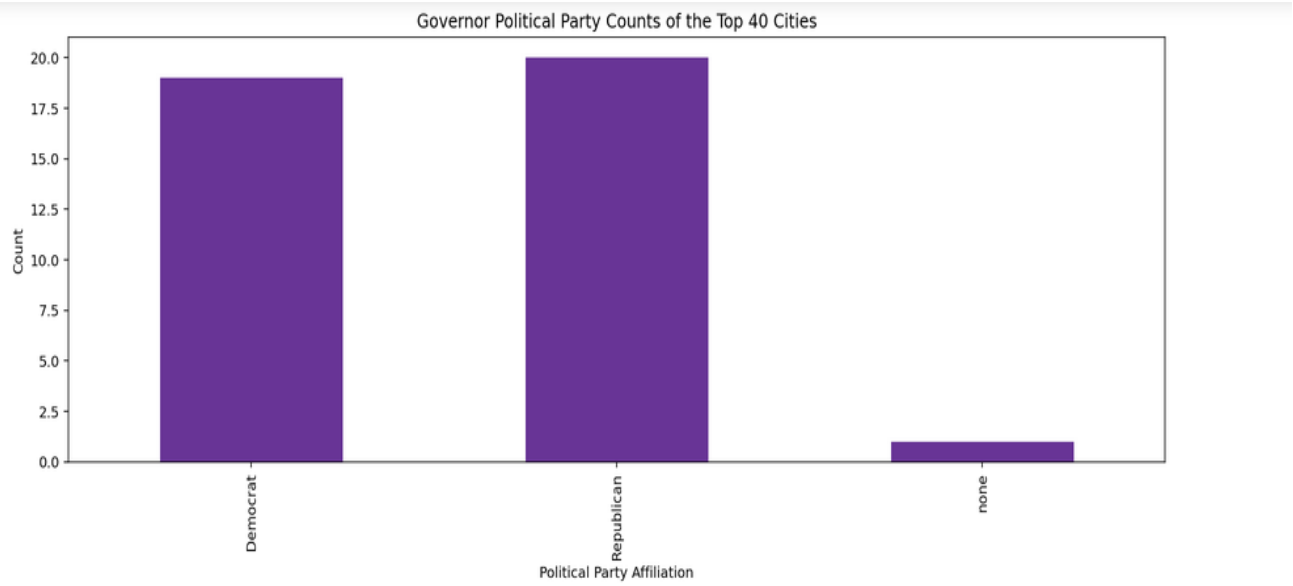
Mayoral Party Counts for Top 40, Mayor Party Counts by Region



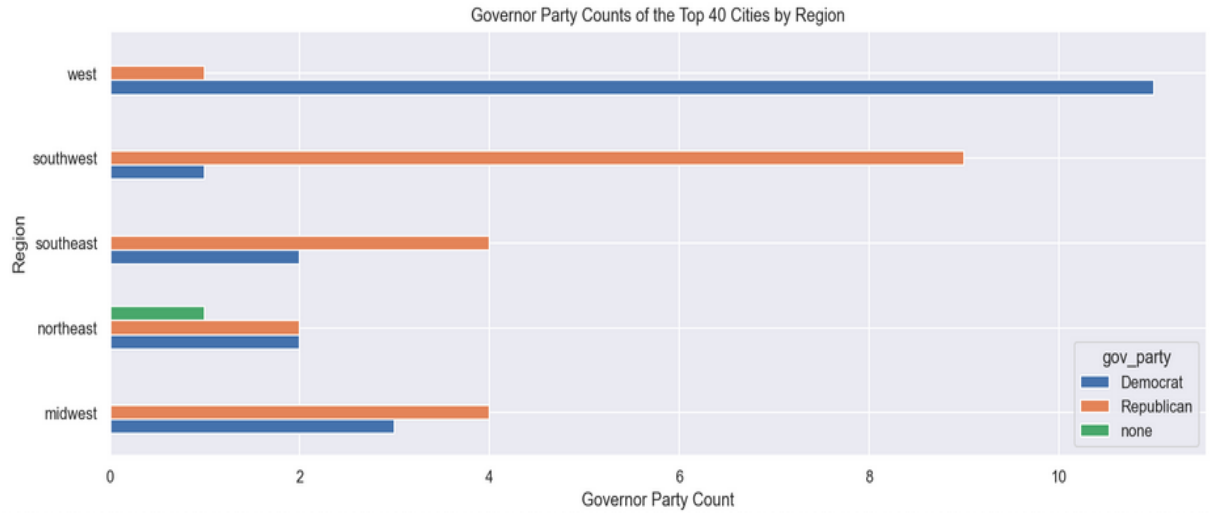
In [115]:
```
mayor_region_plot = top_df.groupby("region")["mayor_party"].value_counts().unstack()
mayor_region_plot.plot.barh(figsize = (15,5))
plt.xlabel("Mayoral Party Count")
plt.ylabel("Region")
plt.title("Mayoral Party Counts of the Top 40 Cities by Region");
```



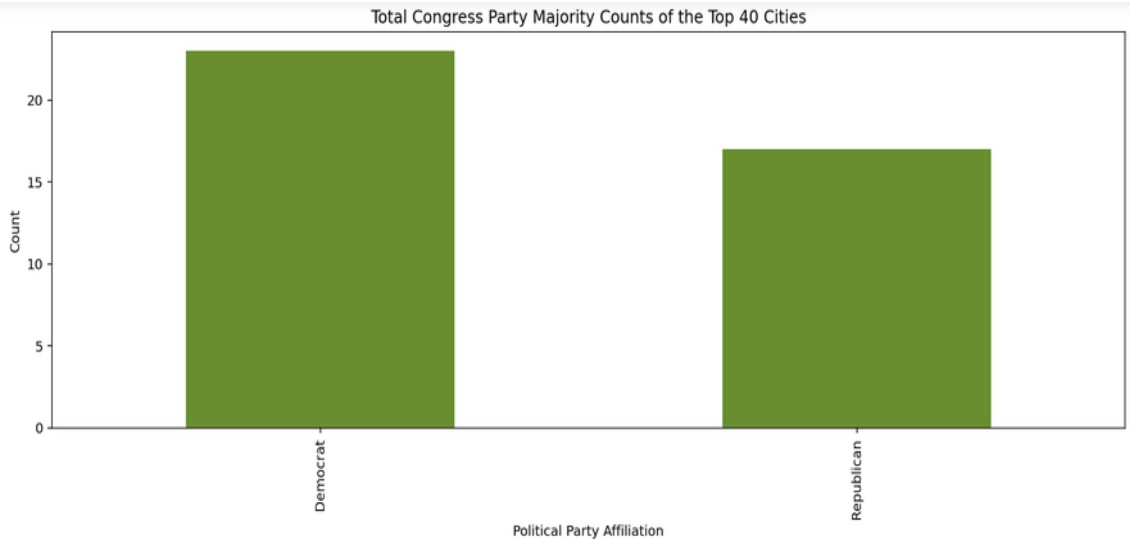
Governor Party Count for Top 40 Cities, Governor Party Counts by Region



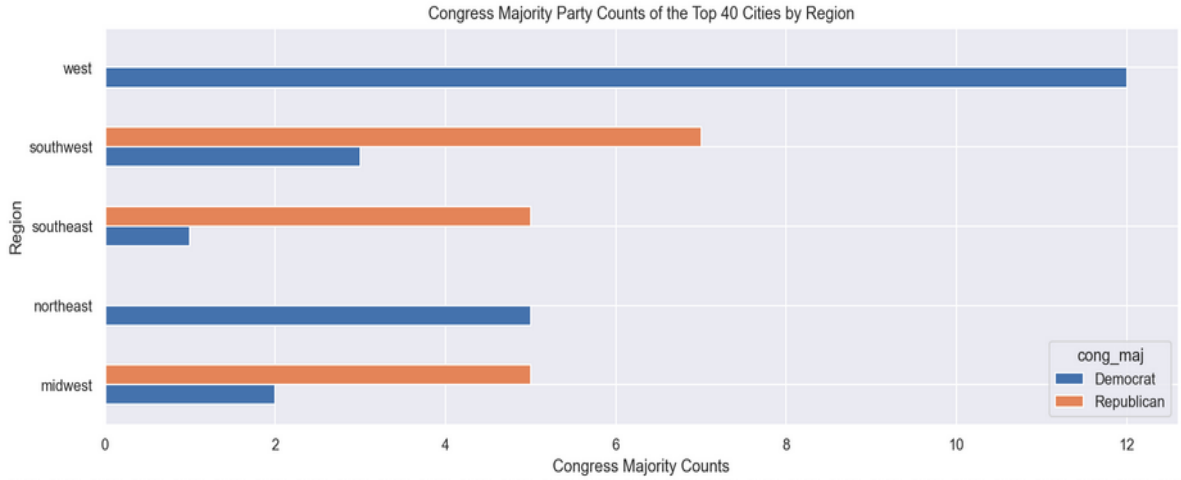

```
n [116]: gov_region_plot = top_df.groupby("region")["gov_party"].value_counts().unstack()
         gov_region_plot.plot.barh(figsize = (15,5))
         plt.xlabel("Governor Party Count")
         plt.ylabel("Region")
         plt.title("Governor Party Counts of the Top 40 Cities by Region");
```



Congress Majority for Top 40 Cities, Congress Majority by Region



In [117]:
```
cong_region_plot = top_df.groupby("region")["cong_maj"].value_counts().unstack()
cong_region_plot.plot.barh(figsize = (15,5))
plt.xlabel("Congress Majority Counts")
plt.ylabel("Region")
plt.title("Congress Majority Party Counts of the Top 40 Cities by Region");
```



Total Political Party Majority Counts of the Top 40 Cities by Region