

Week 10 Pandas Challenge Questions – Tom Purcell – 01/29/2023

1. Is there any missing data? If so, visualize it in a plot.
2. Remove any columns that you may find not useful for your analysis.
3. Plot the top 10 cities with the most accidents.
4. Plot the distribution of the start time.
5. Is the distribution of accidents by hour the same on weekends as on weekdays?
6. What is the distribution of start time on Sundays?
7. Create a geographical plot to show accidents across a map of the United States.
8. Among the top 100 cities in number of accidents, which states do they belong to most frequently?
9. Which months have the most accidents?
10. What is the trend of accidents year over year (decreasing/increasing?)
11. Is the state of New York in the data?
12. Come up with your own question and answer it

1.

City missing 50 entries. Airport\_code missing 3571 entries. Temperature missing 25964. Precipitation missing 208769. Weather Condition missing 26447.

Figure 1 Heatmap of missing values

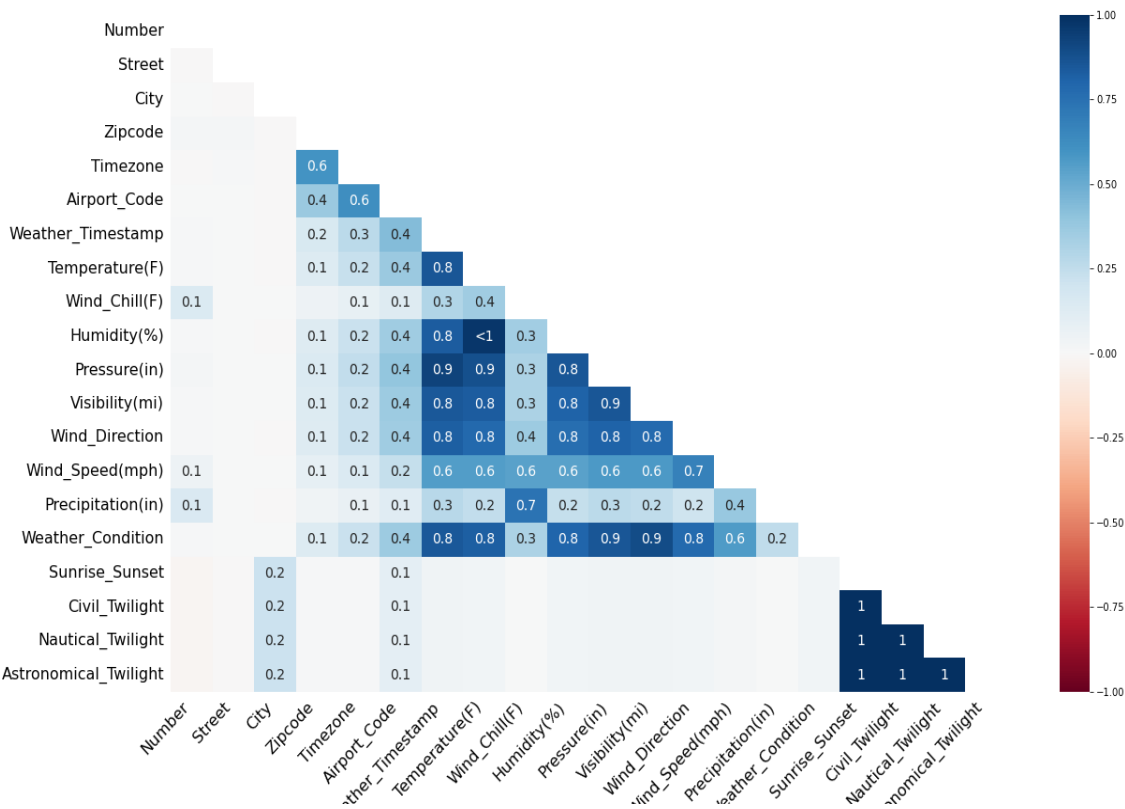
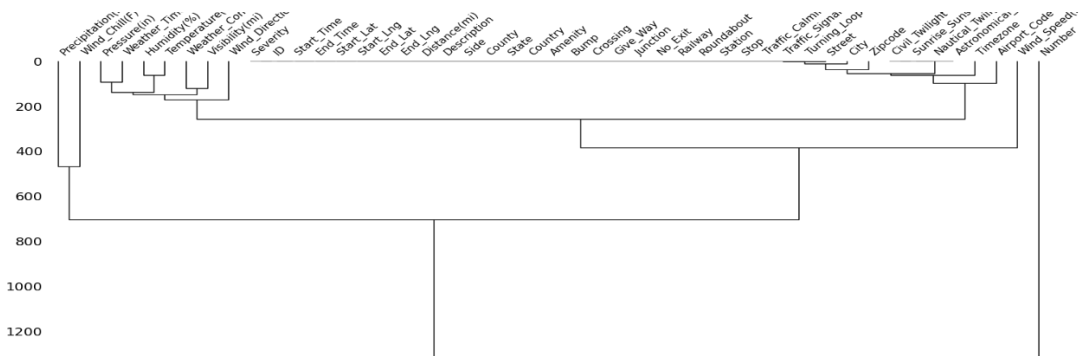


Figure 2 Dendrogram of missing values



2.

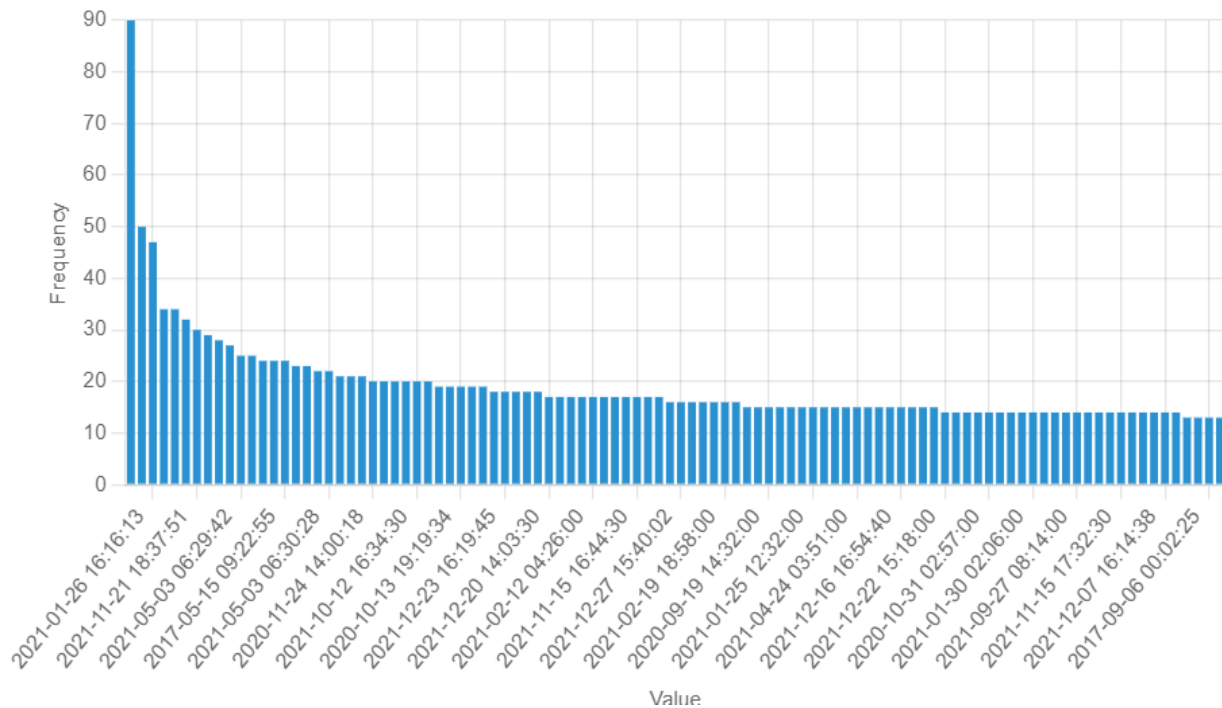
Removed all but 13 columns.

| Column Name       |
|-------------------|
| ID                |
| Severity          |
| Start_Time        |
| City              |
| State             |
| Zipcode           |
| Airport_Code      |
| Temperature(F)    |
| Precipitation(in) |
| Crossing          |
| Railway           |
| Stop              |
| Traffic_Signal    |

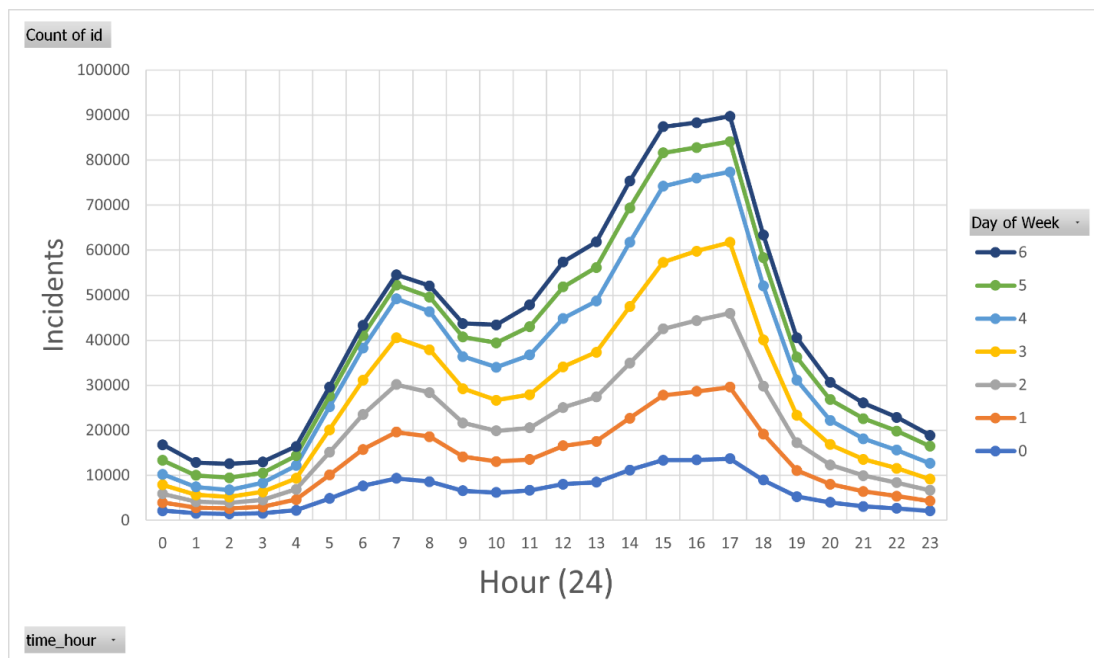
3. Top ten cities - table

| City        | ID count |
|-------------|----------|
| Miami       | 42679    |
| Los Angeles | 27577    |
| Orlando     | 21876    |
| Dallas      | 16863    |
| Houston     | 15710    |
| Charlotte   | 13364    |
| Sacramento  | 12858    |
| San Diego   | 10586    |
| Raleigh     | 9154     |
| Minneapolis | 9137     |

4. Plot the distribution of the start time.

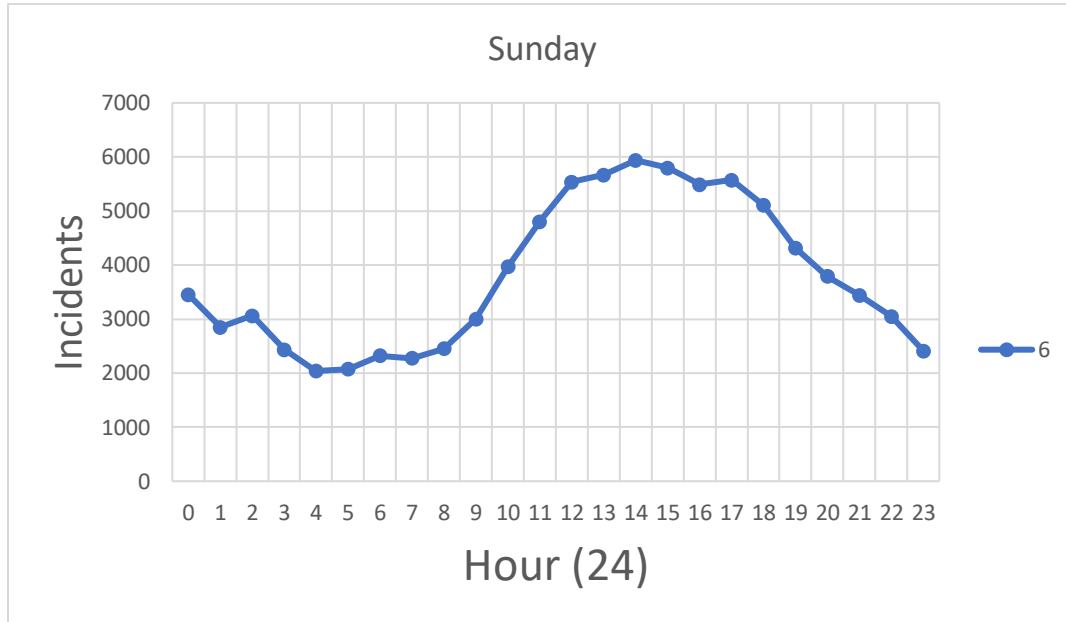


5. Is the distribution of accidents by hour the same on weekends as on weekdays?

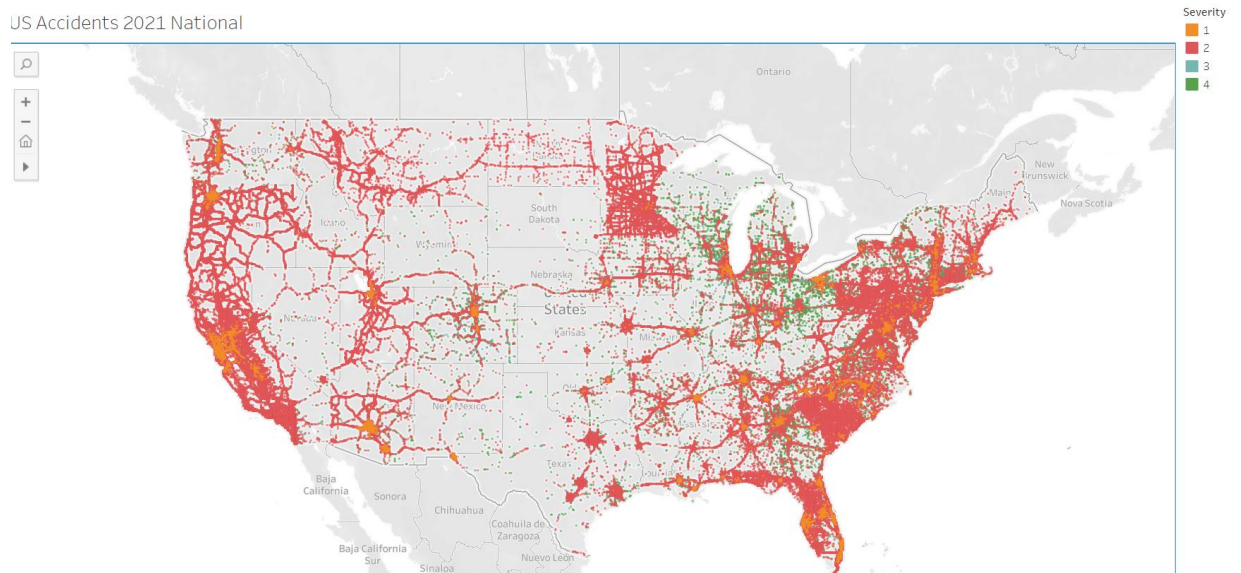


Weekends have a higher rate of accidents by hour than weekdays.

6. What is the distribution of start time on Sundays?



7. Create a geographical plot to show accidents across a map of the United States.



8. Among the top 100 cities in number of accidents, which states do they belong to most frequently?

| State | City            | ID<br>count |
|-------|-----------------|-------------|
| FL    | Miami           | 106896      |
| CA    | Los Angeles     | 68956       |
| FL    | Orlando         | 54654       |
| TX    | Dallas          | 41615       |
| TX    | Houston         | 39366       |
| NC    | Charlotte       | 32998       |
| CA    | Sacramento      | 32559       |
| CA    | San Diego       | 26627       |
| NC    | Raleigh         | 22837       |
| MN    | Minneapolis     | 22768       |
| OR    | Portland        | 20563       |
| TN    | Nashville       | 20160       |
| LA    | Baton Rouge     | 18182       |
| TX    | Austin          | 17928       |
| AZ    | Phoenix         | 16940       |
| MN    | Saint Paul      | 16727       |
| LA    | New Orleans     | 16251       |
| GA    | Atlanta         | 15614       |
| FL    | Jacksonville    | 14472       |
| AZ    | Tucson          | 13799       |
| CA    | San Jose        | 13376       |
| CA    | Riverside       | 12861       |
| CA    | Bakersfield     | 12044       |
| VA    | Richmond        | 11990       |
| CA    | San Bernardino  | 11249       |
| IL    | Chicago         | 11188       |
| FL    | Tampa           | 10984       |
| UT    | Salt Lake City  | 10709       |
| MO    | Kansas City     | 10700       |
| CA    | Anaheim         | 10502       |
| FL    | Hialeah         | 10398       |
| FL    | Fort Lauderdale | 10345       |
| CA    | Ontario         | 9719        |
| NY    | Bronx           | 9453        |
| CA    | Oakland         | 9255        |
| TN    | Chattanooga     | 9205        |
| DC    | Washington      | 9021        |

|    |               |      |
|----|---------------|------|
| CA | Fresno        | 9021 |
| WA | Seattle       | 8896 |
| CA | Long Beach    | 8862 |
| CA | Corona        | 8850 |
| SC | Columbia      | 8664 |
| IN | Indianapolis  | 8328 |
| NY | Rochester     | 8224 |
| CO | Denver        | 7729 |
| PA | Philadelphia  | 7530 |
| CA | Whittier      | 7162 |
| NY | New York      | 7068 |
| MI | Detroit       | 7059 |
| TX | Fort Worth    | 7029 |
| FL | Bradenton     | 7021 |
| FL | Fort Myers    | 6971 |
| CA | San Francisco | 6924 |
| FL | Sarasota      | 6908 |
| FL | Kissimmee     | 6858 |
| MI | Grand Rapids  | 6712 |
| TX | San Antonio   | 6665 |
| SC | Greenville    | 6614 |
|    | Saint         |      |
| FL | Petersburg    | 6199 |
| LA | Shreveport    | 6145 |
| FL | Pensacola     | 6098 |
| NY | Brooklyn      | 6035 |
| MO | Saint Louis   | 5732 |
| CA | Stockton      | 5438 |
| VA | Norfolk       | 5088 |
| CA | Hayward       | 5053 |
| CA | Irvine        | 4990 |
| OH | Dayton        | 4955 |
| FL | North Miami   | 4704 |
| OK | Tulsa         | 4695 |
| CA | Baldwin Park  | 4694 |
| CA | Fontana       | 4575 |
| CA | Tracy         | 4456 |
| CA | Pomona        | 4446 |
| CA | Santa Ana     | 4400 |
| CA | Modesto       | 4299 |
| CA | Pasadena      | 4223 |
| CA | Gardena       | 4166 |
| PA | Pittsburgh    | 4149 |

|    |               |      |
|----|---------------|------|
| FL | Ocala         | 4149 |
| FL | Homestead     | 4116 |
| TN | Memphis       | 4113 |
| MI | Flint         | 4068 |
| MD | Silver Spring | 4041 |
| CA | Sherman Oaks  | 4022 |
| FL | Hollywood     | 3962 |
| PA | Lancaster     | 3923 |
| AL | Birmingham    | 3920 |
| CA | Fremont       | 3823 |
| VA | Fairfax       | 3763 |
| CA | Colton        | 3737 |
| FL | Opa Locka     | 3683 |
| MO | Springfield   | 3647 |
| OH | Cincinnati    | 3628 |
| OR | Salem         | 3611 |
| CA | Orange        | 3593 |
| OH | Columbus      | 3550 |
| UT | Ogden         | 3542 |
| OR | Bend          | 3540 |
| MD | Frederick     | 3524 |

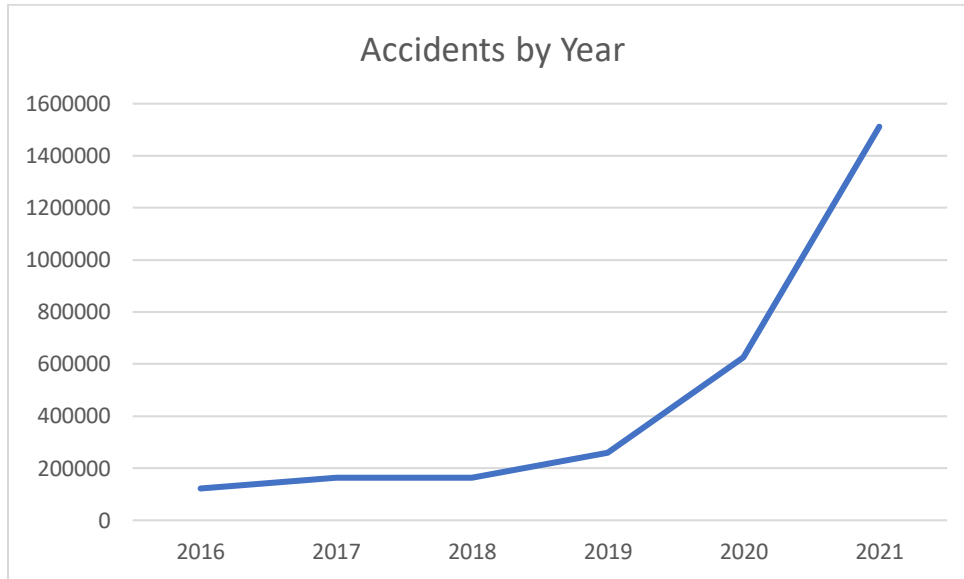
9. Which months have the most accidents?

|           |        |
|-----------|--------|
| December  | 473943 |
| November  | 360696 |
| October   | 299131 |
| September | 241822 |
| June      | 226561 |
| January   | 198365 |
| February  | 194995 |
| May       | 181944 |
| August    | 178670 |
| April     | 171880 |
| July      | 159111 |
| March     | 158224 |



10. What is the trend of accidents year over year (decreasing/increasing?)

***Increasing. Possible the sudden rise in 2018 would be related to reporting accuracy from all 50 states. The weather may be affecting the number increase as well.***

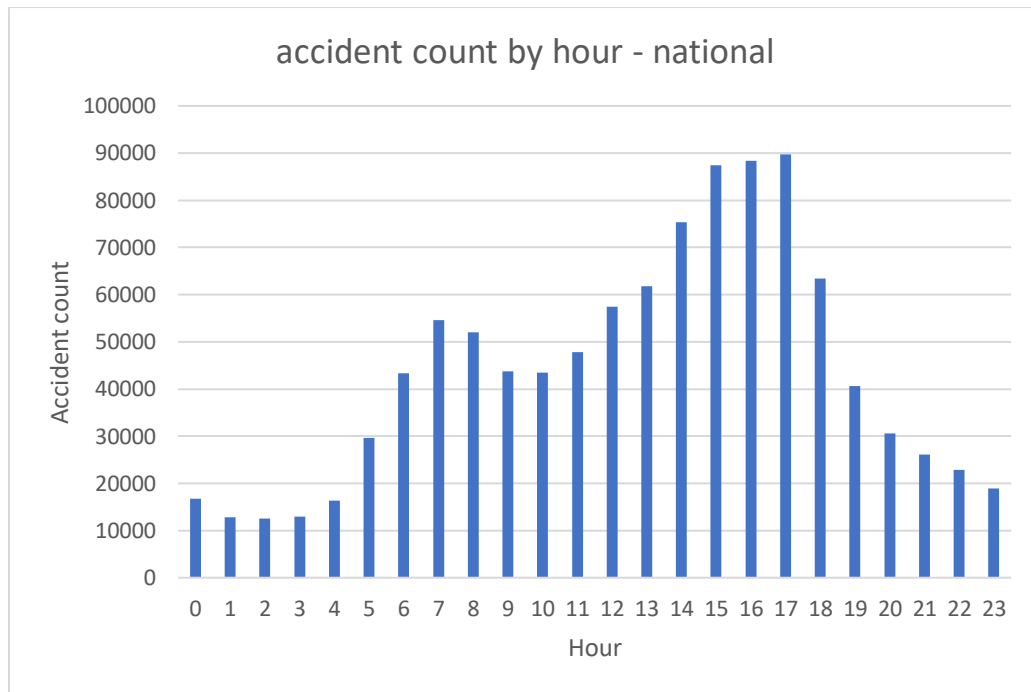


| year | count   |
|------|---------|
| 2016 | 122024  |
| 2017 | 163918  |
| 2018 | 163176  |
| 2019 | 258615  |
| 2020 | 625864  |
| 2021 | 1511745 |

11. Is the state of New York in the data? Yes.

12. Come up with your own question and answer it.

a. What is the accident count by hour nationally?



b. Display zipcodes that have the highest occurring # of accidents sorted by # accidents DESC?

