

## Week 13 - K-means Clustering

Devin Wilson

4/3/2023

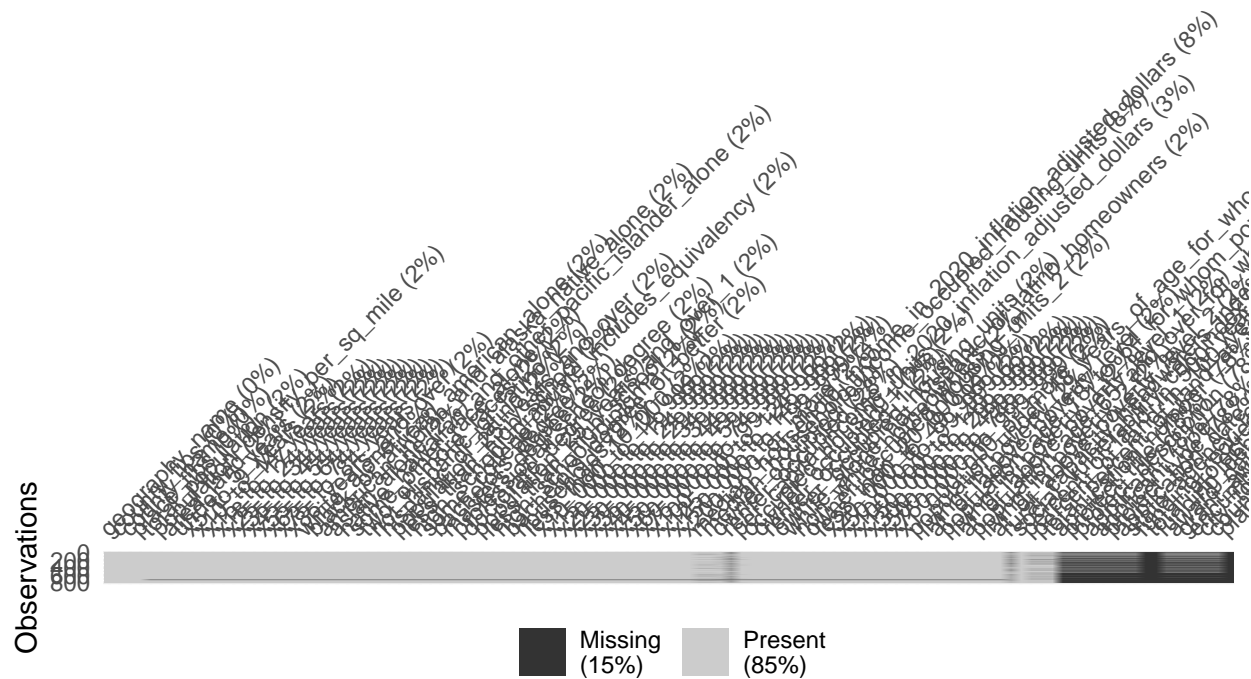
```
knitr::opts_chunk$set(root.dir='Users/devinwilson/Documents/EDF/Flood_Analysis_PLCY_698') # change for
```

## #Running Libraries and Reading in the Data

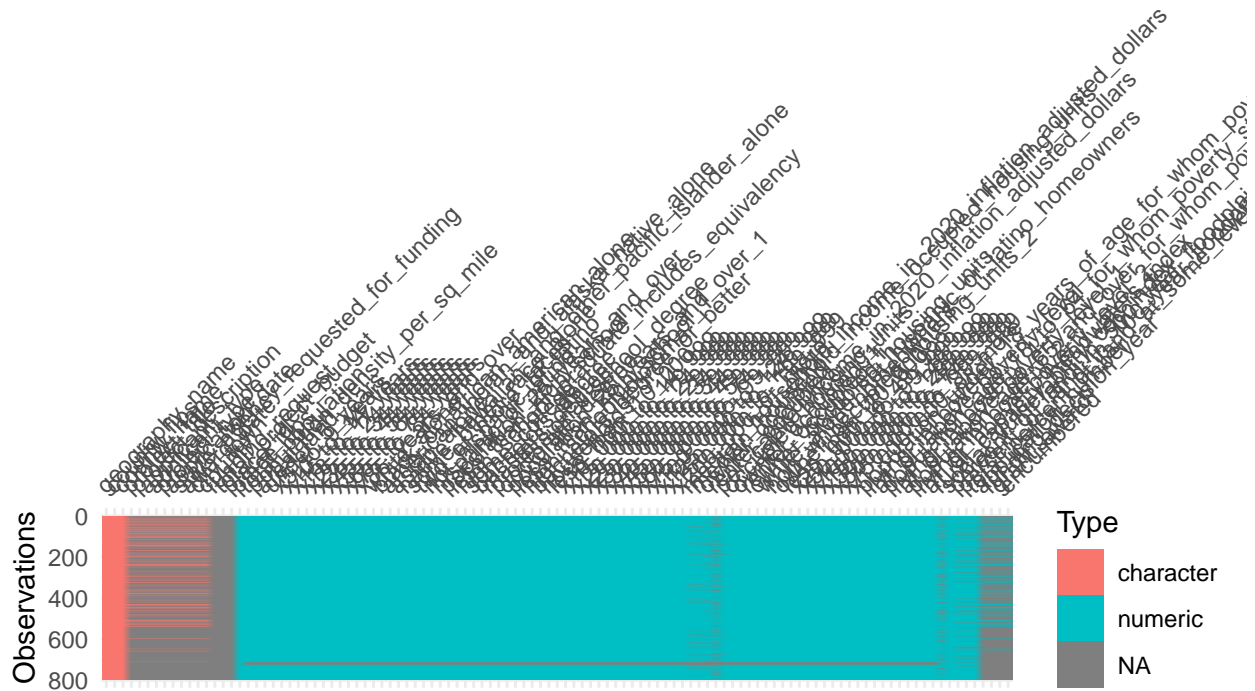
## EDA

```
#install Packages
library(visdat)
library(naniar)

# visualize missingness
vis_miss(data)
```



```
vis_dat(data)
```



```
data %>% summarise(across(everything(), ~ mean(is.na(.))))
```

```
## # A tibble: 1 x 109
##   geography_name count~1 count~2 place~3 total~4 popul~5 area_~6 under~7 x5_to~8
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      0      0      0      0 0.0163 0.0163 0.0163 0.0163 0.0163
## # ... with 100 more variables: x10_to_14_years <dbl>, x15_to_17_years <dbl>,
## #   x18_to_24_years <dbl>, x25_to_34_years <dbl>, x35_to_44_years <dbl>,
## #   x45_to_54_years <dbl>, x55_to_64_years <dbl>, x65_to_74_years <dbl>,
## #   x75_to_84_years <dbl>, x85_years_and_over <dbl>, white_alone <dbl>,
## #   black_or_african_american_alone <dbl>,
## #   american_indian_and_alaska_native_alone <dbl>, asian_alone <dbl>,
## #   native_hawaiian_and_other_pacific_islander_alone <dbl>, ...
```

```
#Looking at summary statistics
```

```
glimpse(data)
```

```
## Rows: 796
## Columns: 109
## $ geography_name <chr> ~
## $ county_fips <chr> ~
## $ county_name <chr> ~
## $ place_fips <dbl> ~
## $ total_population <dbl> ~
## $ population_density_per_sq_mile <dbl> ~
## $ area_land <dbl> ~
## $ under_5_years <dbl> ~
## $ x5_to_9_years <dbl> ~
## $ x10_to_14_years <dbl> ~
## $ x15_to_17_years <dbl> ~
## $ x18_to_24_years <dbl> ~
## $ x25_to_34_years <dbl> ~
## $ x35_to_44_years <dbl> ~
```

```

## $ x45_to_54_years <dbl> ~
## $ x55_to_64_years <dbl> ~
## $ x65_to_74_years <dbl> ~
## $ x75_to_84_years <dbl> ~
## $ x85_years_and_over <dbl> ~
## $ white_alone <dbl> ~
## $ black_or_african_american_alone <dbl> ~
## $ american_indian_and_alaska_native_alone <dbl> ~
## $ asian_alone <dbl> ~
## $ native_hawaiian_and_other_pacific_islander_alone <dbl> ~
## $ some_other_race_alone <dbl> ~
## $ two_or_more_races <dbl> ~
## $ not_hispanic_or_latino <dbl> ~
## $ hispanic_or_latino <dbl> ~
## $ population_25_years_and_over <dbl> ~
## $ less_than_high_school <dbl> ~
## $ high_school_graduate_includes_equivalency <dbl> ~
## $ some_college <dbl> ~
## $ bachelors_degree <dbl> ~
## $ masters_degree <dbl> ~
## $ professional_school_degree <dbl> ~
## $ doctorate_degree <dbl> ~
## $ population_25_years_and_over_1 <dbl> ~
## $ less_than_high_school_1 <dbl> ~
## $ high_school_diploma <dbl> ~
## $ bachelors_degree_or_better <dbl> ~
## $ households <dbl> ~
## $ less_than_10_000 <dbl> ~
## $ x10_000_to_14_999 <dbl> ~
## $ x15_000_to_19_999 <dbl> ~
## $ x20_000_to_24_999 <dbl> ~
## $ x25_000_to_29_999 <dbl> ~
## $ x30_000_to_34_999 <dbl> ~
## $ x35_000_to_39_999 <dbl> ~
## $ x40_000_to_44_999 <dbl> ~
## $ x45_000_to_49_999 <dbl> ~
## $ x50_000_to_59_999 <dbl> ~
## $ x60_000_to_74_999 <dbl> ~
## $ x75_000_to_99_999 <dbl> ~
## $ x100_000_to_124_999 <dbl> ~
## $ x125_000_to_149_999 <dbl> ~
## $ x150_000_to_199_999 <dbl> ~
## $ x200_000_or_more <dbl> ~
## $ median_household_income_in_2020_inflation_adjusted_dollars <dbl> ~
## $ median_household_income_occupied_housing_units <dbl> ~
## $ owner_occupied <dbl> ~
## $ renter_occupied <dbl> ~
## $ per_capita_income_in_2020_inflation_adjusted_dollars <dbl> ~
## $ occupied_housing_units <dbl> ~
## $ owner_occupied_1 <dbl> ~
## $ renter_occupied_1 <dbl> ~
## $ owner_occupied_housing_units <dbl> ~
## $ white_alone_not_hispanic_or_latino_homeowners <dbl> ~
## $ non_white_homeowners <dbl> ~

```

```
## $ owner_occupied_housing_units_2 <dbl> ~
## $ less_than_20_000 <dbl> ~
## $ x20_000_to_49_999 <dbl> ~
## $ x50_000_to_99_999 <dbl> ~
## $ x100_000_to_149_999 <dbl> ~
## $ x150_000_to_299_999 <dbl> ~
## $ x300_000_to_499_999 <dbl> ~
## $ x500_000_to_749_999 <dbl> ~
## $ x750_000_to_999_999 <dbl> ~
## $ x1_000_000_or_more <dbl> ~
## $ population_under_18_years_of_age_for_whom_poverty_status_is_determined <dbl> ~
## $ living_in_poverty <dbl> ~
## $ at_or_above_poverty_level <dbl> ~
## $ population_age_18_to_64_for_whom_poverty_status_is_determined <dbl> ~
## $ living_in_poverty_1 <dbl> ~
## $ at_or_above_poverty_level_1 <dbl> ~
## $ population_age_65_and_over_for_whom_poverty_status_is_determined <dbl> ~
## $ living_in_poverty_2 <dbl> ~
## $ at_or_above_poverty_level_2 <dbl> ~
## $ rural_capacity_headwaters_index <dbl> ~
## $ social_vulnerability_cdc_index <dbl> ~
## $ percent_of_land_in_500_year_floodplain <dbl> ~
## $ percent_of_land_in_100_year_floodplain <dbl> ~
## $ total_percent_of_land_at_some_level_of_flood_risk <dbl> ~
## $ application_number <dbl> ~
## $ project_title <chr> ~
## $ project_description <chr> ~
## $ applicant <chr> ~
## $ applicant_type <chr> ~
## $ project_type <chr> ~
## $ application_date <chr> ~
## $ state_agency_requested_for_funding <chr> ~
## $ funding_request <lgl> ~
## $ total_project_budget <lgl> ~
## $ funding_decision_year <dbl> ~
## $ award_date <chr> ~
## $ grant_award <dbl> ~
## $ encumbered <dbl> ~
## $ city_town <chr> ~
## $ county <chr> ~
## $ place_id <lgl> ~
```

```
summary(data)
```

```
## geography_name county_fips county_name place_fips
## Length:796 Length:796 Length:796 Min. :3700160
## Class :character Class :character Class :character 1st Qu.:3718315
## Mode :character Mode :character Mode :character Median :3736910
## Mean :3737286
## 3rd Qu.:3757005
## Max. :3776220
##
## total_population population_density_per_sq_mile area_land
## Min. : 0 Min. : 0.0 Min. : 0.11
## 1st Qu.: 535 1st Qu.: 319.9 1st Qu.: 1.27
```

```

## Median : 1451      Median : 674.8              Median : 2.76
## Mean : 8212      Mean : 797.7              Mean : 6.51
## 3rd Qu.: 4102      3rd Qu.:1131.5              3rd Qu.: 6.15
## Max. :873570      Max. :5240.2              Max. :308.29
## NA's :13      NA's :13              NA's :13
## under_5_years      x5_to_9_years      x10_to_14_years      x15_to_17_years
## Min. : 0.0      Min. : 0.0      Min. : 0.0      Min. : 0.0
## 1st Qu.: 18.0      1st Qu.: 21.0      1st Qu.: 21.5      1st Qu.: 14.0
## Median : 72.0      Median : 82.0      Median : 91.0      Median : 51.0
## Mean : 505.7      Mean : 510.8      Mean : 525.5      Mean : 309.9
## 3rd Qu.: 229.5      3rd Qu.: 239.0      3rd Qu.: 249.0      3rd Qu.: 158.0
## Max. :58918.0      Max. :56624.0      Max. :56281.0      Max. :32565.0
## NA's :13      NA's :13      NA's :13      NA's :13
## x18_to_24_years      x25_to_34_years      x35_to_44_years      x45_to_54_years
## Min. : 0.0      Min. : 0      Min. : 0      Min. : 0
## 1st Qu.: 31.0      1st Qu.: 55      1st Qu.: 49      1st Qu.: 63
## Median : 99.0      Median : 170      Median : 163      Median : 196
## Mean : 888.9      Mean : 1203      Mean : 1058      Mean : 1034
## 3rd Qu.: 342.0      3rd Qu.: 495      3rd Qu.: 502      3rd Qu.: 519
## Max. :82299.0      Max. :159567      Max. :129349      Max. :112992
## NA's :13      NA's :13      NA's :13      NA's :13
## x55_to_64_years      x65_to_74_years      x75_to_84_years      x85_years_and_over
## Min. : 0.0      Min. : 0.0      Min. : 0.0      Min. : 0.0
## 1st Qu.: 69.5      1st Qu.: 56.0      1st Qu.: 27.0      1st Qu.: 6.0
## Median : 200.0      Median : 182.0      Median : 86.0      Median : 30.0
## Mean : 950.9      Mean : 716.8      Mean : 352.7      Mean : 155.1
## 3rd Qu.: 556.0      3rd Qu.: 492.0      3rd Qu.: 254.0      3rd Qu.: 101.0
## Max. :94216.0      Max. :56165.0      Max. :24004.0      Max. :10590.0
## NA's :13      NA's :13      NA's :13      NA's :13
## white_alone      black_or_african_american_alone
## Min. : 0.0      Min. : 0
## 1st Qu.: 309.5      1st Qu.: 17
## Median : 1091.0      Median : 161
## Mean : 5017.5      Mean : 2175
## 3rd Qu.: 3132.5      3rd Qu.: 708
## Max. :407684.0      Max. :309837
## NA's :13      NA's :13
## american_indian_and_alaska_native_alone      asian_alone
## Min. : 0.00      Min. : 0.0
## 1st Qu.: 0.00      1st Qu.: 0.0
## Median : 0.00      Median : 0.0
## Mean : 57.69      Mean : 339.4
## 3rd Qu.: 23.00      3rd Qu.: 38.0
## Max. :3468.00      Max. :57889.0
## NA's :13      NA's :13
## native_hawaiian_and_other_pacific_islander_alone      some_other_race_alone
## Min. : 0.000      Min. : 0.0
## 1st Qu.: 0.000      1st Qu.: 0.0
## Median : 0.000      Median : 8.0
## Mean : 6.146      Mean : 286.5
## 3rd Qu.: 0.000      3rd Qu.: 81.0
## Max. :1014.000      Max. :58164.0
## NA's :13      NA's :13
## two_or_more_races      not_hispanic_or_latino      hispanic_or_latino

```

```

## Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 1.0 1st Qu.: 495.5 1st Qu.: 7.0
## Median : 29.0 Median : 1361.0 Median : 65.0
## Mean : 329.3 Mean : 7353.2 Mean : 858.5
## 3rd Qu.: 148.0 3rd Qu.: 3780.0 3rd Qu.: 315.0
## Max. :36179.0 Max. :746265.0 Max. :127305.0
## NA's :13 NA's :13 NA's :13
## population_25_years_and_over less_than_high_school
## Min. : 0 Min. : 0.0
## 1st Qu.: 393 1st Qu.: 40.5
## Median : 1031 Median : 125.0
## Mean : 5471 Mean : 571.5
## 3rd Qu.: 2968 3rd Qu.: 342.5
## Max. :586883 Max. :62042.0
## NA's :13 NA's :13
## high_school_graduate_includes_equivalency some_college
## Min. : 0 Min. : 0.0
## 1st Qu.: 115 1st Qu.: 117.5
## Median : 308 Median : 339.0
## Mean : 1204 Mean : 1621.7
## 3rd Qu.: 774 3rd Qu.: 961.5
## Max. :98608 Max. :163444.0
## NA's :13 NA's :13
## bachelors_degree masters_degree professional_school_degree
## Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 33.5 1st Qu.: 9.0 1st Qu.: 0.0
## Median : 139.0 Median : 47.0 Median : 5.0
## Mean : 1299.9 Mean : 552.9 Mean : 128.6
## 3rd Qu.: 505.5 3rd Qu.: 197.5 3rd Qu.: 40.0
## Max. :171566.0 Max. :68365.0 Max. :16246.0
## NA's :13 NA's :13 NA's :13
## doctorate_degree population_25_years_and_over_1 less_than_high_school_1
## Min. : 0.00 Min. : 0 Min. : 0.0
## 1st Qu.: 0.00 1st Qu.: 393 1st Qu.: 40.5
## Median : 3.00 Median : 1031 Median : 125.0
## Mean : 92.14 Mean : 5471 Mean : 571.5
## 3rd Qu.: 26.50 3rd Qu.: 2968 3rd Qu.: 342.5
## Max. :8701.00 Max. :586883 Max. :62042.0
## NA's :13 NA's :13 NA's :13
## high_school_diploma bachelors_degree_or_better households
## Min. : 0.0 Min. : 0.0 Min. : 0
## 1st Qu.: 243.5 1st Qu.: 51.5 1st Qu.: 214
## Median : 641.0 Median : 207.0 Median : 609
## Mean : 2825.8 Mean : 2073.5 Mean : 3198
## 3rd Qu.: 1708.5 3rd Qu.: 821.0 3rd Qu.: 1572
## Max. :262052.0 Max. :262789.0 Max. :338985
## NA's :13 NA's :13 NA's :13
## less_than_10_000 x10_000_to_14_999 x15_000_to_19_999 x20_000_to_24_999
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 8.0 1st Qu.: 7.0 1st Qu.: 8.0 1st Qu.: 8.0
## Median : 39.0 Median : 29.0 Median : 32.0 Median : 28.0
## Mean : 212.1 Mean : 153.2 Mean : 148.7 Mean : 155.5
## 3rd Qu.: 111.0 3rd Qu.: 90.5 3rd Qu.: 89.0 3rd Qu.: 85.0
## Max. :16173.0 Max. :11650.0 Max. :11812.0 Max. :12715.0

```

```

## NA's :13      NA's :13      NA's :13      NA's :13
## x25_000_to_29_999 x30_000_to_34_999 x35_000_to_39_999 x40_000_to_44_999
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 8.0 1st Qu.: 7.0 1st Qu.: 6.0 1st Qu.: 4.5
## Median : 29.0 Median : 28.0 Median : 25.0 Median : 22.0
## Mean : 153.3 Mean : 159.8 Mean : 150.9 Mean : 147.6
## 3rd Qu.: 83.5 3rd Qu.: 79.0 3rd Qu.: 74.0 3rd Qu.: 76.5
## Max. :13865.0 Max. :16902.0 Max. :15431.0 Max. :16143.0
## NA's :13      NA's :13      NA's :13      NA's :13
## x45_000_to_49_999 x50_000_to_59_999 x60_000_to_74_999 x75_000_to_99_999
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 4.0 1st Qu.: 14.0 1st Qu.: 17.0 1st Qu.: 19.0
## Median : 20.0 Median : 46.0 Median : 60.0 Median : 64.0
## Mean : 128.7 Mean : 252.6 Mean : 307.9 Mean : 391.2
## 3rd Qu.: 60.5 3rd Qu.: 132.0 3rd Qu.: 165.0 3rd Qu.: 202.5
## Max. :13134.0 Max. :28467.0 Max. :33903.0 Max. :43066.0
## NA's :13      NA's :13      NA's :13      NA's :13
## x100_000_to_124_999 x125_000_to_149_999 x150_000_to_199_999 x200_000_or_more
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 8.0 1st Qu.: 1.5 1st Qu.: 0.0 1st Qu.: 0.0
## Median : 37.0 Median : 17.0 Median : 13.0 Median : 10.0
## Mean : 267.3 Mean : 167.8 Mean : 184.1 Mean : 217.9
## 3rd Qu.: 145.5 3rd Qu.: 76.5 3rd Qu.: 70.0 3rd Qu.: 55.5
## Max. :30997.0 Max. :19679.0 Max. :22236.0 Max. :32812.0
## NA's :13      NA's :13      NA's :13      NA's :13
## median_household_income_in_2020_inflation_adjusted_dollars
## Min. : 13889
## 1st Qu.: 36613
## Median : 47432
## Mean : 53432
## 3rd Qu.: 64125
## Max. :209950
## NA's :63
## median_household_income_occupied_housing_units owner_occupied
## Min. : 13889 Min. : 13492
## 1st Qu.: 36613 1st Qu.: 47586
## Median : 47432 Median : 59315
## Mean : 53432 Mean : 64139
## 3rd Qu.: 64125 3rd Qu.: 74802
## Max. :209950 Max. :210515
## NA's :63 NA's :68
## renter_occupied per_capita_income_in_2020_inflation_adjusted_dollars
## Min. : 2499 Min. : 5987
## 1st Qu.: 22372 1st Qu.: 20402
## Median : 30803 Median : 25050
## Mean : 35482 Mean : 28081
## 3rd Qu.: 41898 3rd Qu.: 33086
## Max. :250001 Max. :145777
## NA's :225 NA's :21
## occupied_housing_units owner_occupied_1 renter_occupied_1
## Min. : 0 Min. : 0.0 Min. : 0.0
## 1st Qu.: 214 1st Qu.: 145.5 1st Qu.: 50.5
## Median : 609 Median : 393.0 Median : 168.0
## Mean : 3198 Mean : 1839.7 Mean : 1358.8

```

```

## 3rd Qu.: 1572          3rd Qu.: 1119.5    3rd Qu.: 504.0
## Max. :338985          Max. :179054.0    Max. :159931.0
## NA's :13              NA's :13          NA's :13
## owner_occupied_housing_units white_alone_not_hispanic_or_latino_homeowners
## Min. : 0.0            Min. : 0
## 1st Qu.: 145.5        1st Qu.: 98
## Median : 393.0        Median : 290
## Mean : 1839.7          Mean : 1315
## 3rd Qu.: 1119.5        3rd Qu.: 923
## Max. :179054.0        Max. :102643
## NA's :13              NA's :13
## non_white_homeowners owner_occupied_housing_units_2 less_than_20_000
## Min. : 0              Min. : 0.0          Min. : 0.00
## 1st Qu.: 12           1st Qu.: 145.5        1st Qu.: 0.00
## Median : 66           Median : 393.0        Median : 9.00
## Mean : 525            Mean : 1839.7        Mean : 39.57
## 3rd Qu.: 196          3rd Qu.: 1119.5        3rd Qu.: 33.50
## Max. :76411           Max. :179054.0        Max. :2388.00
## NA's :13              NA's :13          NA's :13
## x20_000_to_49_999 x50_000_to_99_999 x100_000_to_149_999 x150_000_to_299_999
## Min. : 0.0            Min. : 0.0            Min. : 0.0            Min. : 0.0
## 1st Qu.: 1.0           1st Qu.: 19.0          1st Qu.: 17.5          1st Qu.: 24.0
## Median : 13.0           Median : 60.0           Median : 60.0           Median : 101.0
## Mean : 45.4             Mean : 215.3           Mean : 293.4           Mean : 715.3
## 3rd Qu.: 39.0           3rd Qu.: 147.0          3rd Qu.: 182.0          3rd Qu.: 425.0
## Max. :2026.0            Max. :12628.0          Max. :25722.0          Max. :71187.0
## NA's :13               NA's :13              NA's :13              NA's :13
## x300_000_to_499_999 x500_000_to_749_999 x750_000_to_999_999 x1_000_000_or_more
## Min. : 0.0            Min. : 0.0            Min. : 0.0            Min. : 0.00
## 1st Qu.: 0.0           1st Qu.: 0.0           1st Qu.: 0.0           1st Qu.: 0.00
## Median : 19.0           Median : 0.0           Median : 0.0           Median : 0.00
## Mean : 362.8            Mean : 104.9           Mean : 35.9            Mean : 27.09
## 3rd Qu.: 125.0          3rd Qu.: 22.0          3rd Qu.: 3.0           3rd Qu.: 3.00
## Max. :37736.0           Max. :14767.0          Max. :6588.0           Max. :6444.00
## NA's :13               NA's :13              NA's :13              NA's :13
## population_under_18_years_of_age_for_whom_poverty_status_is_determined
## Min. : 0.0
## 1st Qu.: 92.0
## Median : 291.0
## Mean : 1821.1
## 3rd Qu.: 847.5
## Max. :200673.0
## NA's :13
## living_in_poverty at_or_above_poverty_level
## Min. : 0.0            Min. : 0.0
## 1st Qu.: 6.0           1st Qu.: 63.5
## Median : 53.0           Median : 206.0
## Mean : 378.7            Mean : 1442.5
## 3rd Qu.: 191.0          3rd Qu.: 678.0
## Max. :36185.0           Max. :164488.0
## NA's :13              NA's :13
## population_age_18_to_64_for_whom_poverty_status_is_determined
## Min. : 0.0
## 1st Qu.: 299.5

```



```

## Median : 814.0
## Mean : 4901.4
## 3rd Qu.: 2256.0
## Max. :569328.0
## NA's :13
## living_in_poverty_1 at_or_above_poverty_level_1
## Min. : 0.0 Min. : 0
## 1st Qu.: 34.0 1st Qu.: 240
## Median : 116.0 Median : 692
## Mean : 668.3 Mean : 4233
## 3rd Qu.: 329.0 3rd Qu.: 1950
## Max. :58906.0 Max. :510422
## NA's :13 NA's :13
## population_age_65_and_over_for_whom_poverty_status_is_determined
## Min. : 0.0
## 1st Qu.: 98.5
## Median : 302.0
## Mean : 1180.9
## 3rd Qu.: 785.5
## Max. :89182.0
## NA's :13
## living_in_poverty_2 at_or_above_poverty_level_2
## Min. : 0.0 Min. : 0.0
## 1st Qu.: 5.0 1st Qu.: 83.0
## Median : 24.0 Median : 270.0
## Mean : 110.9 Mean : 1070.0
## 3rd Qu.: 72.0 3rd Qu.: 694.5
## Max. :7530.0 Max. :81652.0
## NA's :13 NA's :13
## rural_capacity_headwaters_index social_vulnerability_cdc_index
## Min. : 4.00 Min. :0.0100
## 1st Qu.:53.00 1st Qu.:0.3500
## Median :58.00 Median :0.6000
## Mean :57.78 Mean :0.5595
## 3rd Qu.:69.00 3rd Qu.:0.7500
## Max. :95.00 Max. :0.9800
## NA's :195
## percent_of_land_in_500_year_floodplain percent_of_land_in_100_year_floodplain
## Min. : 0.000 Min. : 0.00
## 1st Qu.: 0.000 1st Qu.: 3.38
## Median : 0.510 Median : 7.66
## Mean : 2.311 Mean : 15.74
## 3rd Qu.: 1.860 3rd Qu.: 17.00
## Max. :48.430 Max. :100.00
## NA's :93 NA's :93
## total_percent_of_land_at_some_level_of_flood_risk application_number
## Min. : 0.000 Min. : 1.00
## 1st Qu.: 3.655 1st Qu.: 8.00
## Median : 8.690 Median :13.00
## Mean : 18.047 Mean :13.09
## 3rd Qu.: 19.455 3rd Qu.:20.00
## Max. :100.000 Max. :24.00
## NA's :93 NA's :620
## project_title project_description applicant applicant_type

```

```
## Length:796      Length:796      Length:796      Length:796
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
## project_type      application_date  state_agency_requested_for_funding
## Length:796      Length:796      Length:796
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
##
## funding_request total_project_budget funding_decision_year award_date
## Mode:logical    Mode:logical      Min. :2022      Length:796
## NA's:796        NA's:796          1st Qu.:2022    Class :character
##                                     Median :2022      Mode :character
##                                     Mean  :2022
##                                     3rd Qu.:2022
##                                     Max. :2022
##                                     NA's :620
## grant_award      encumbered      city_town      county
## Min. : 90000      Min. : 90000      Length:796      Length:796
## 1st Qu.:175000    1st Qu.:175000    Class :character Class :character
## Median :249880    Median :249880    Mode :character Mode :character
## Mean :216198      Mean :216198
## 3rd Qu.:250000    3rd Qu.:250000
## Max. :250000      Max. :250000
## NA's :620        NA's :620
## place_id
## Mode:logical
## NA's:796
##
##
##
##
```

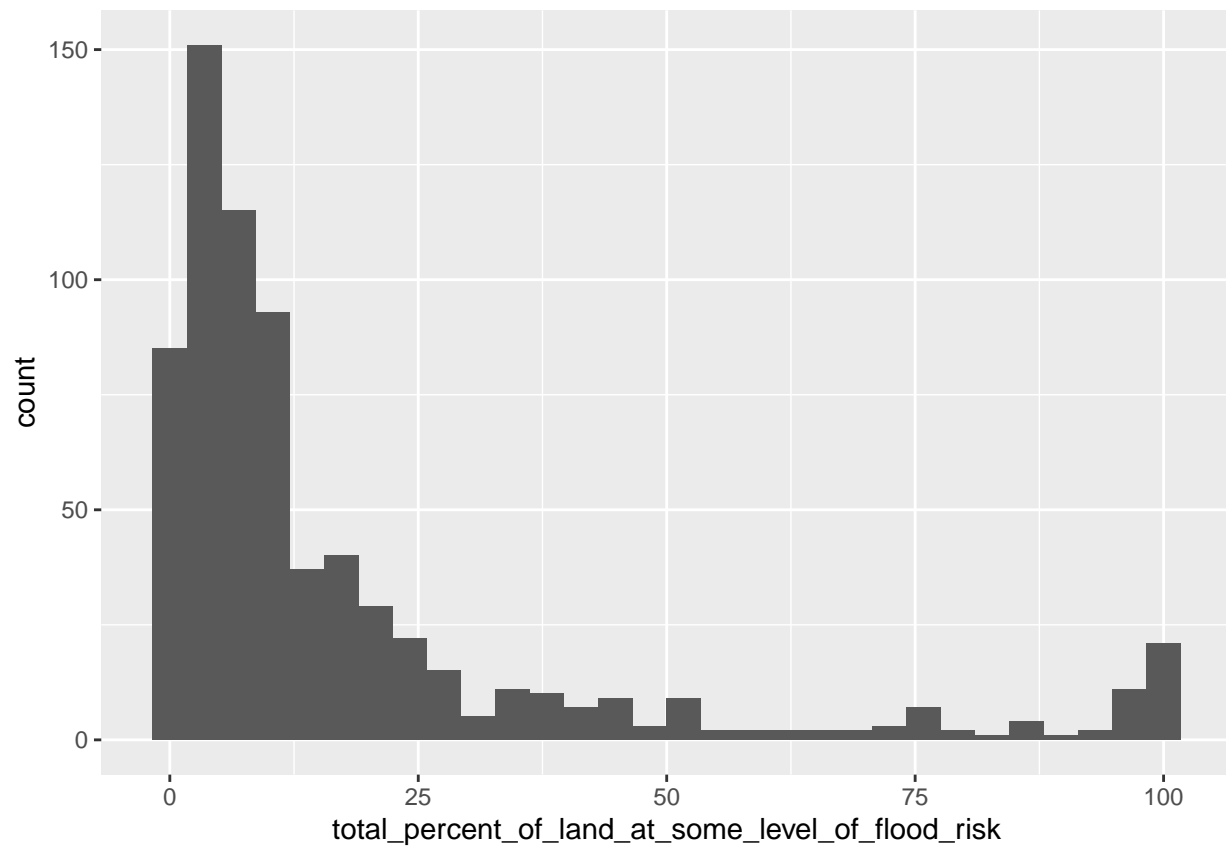
*#Exploring the Data through some plots*

*#Histogram*

```
ggplot(data, aes(x=total_percent_of_land_at_some_level_of_flood_risk)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

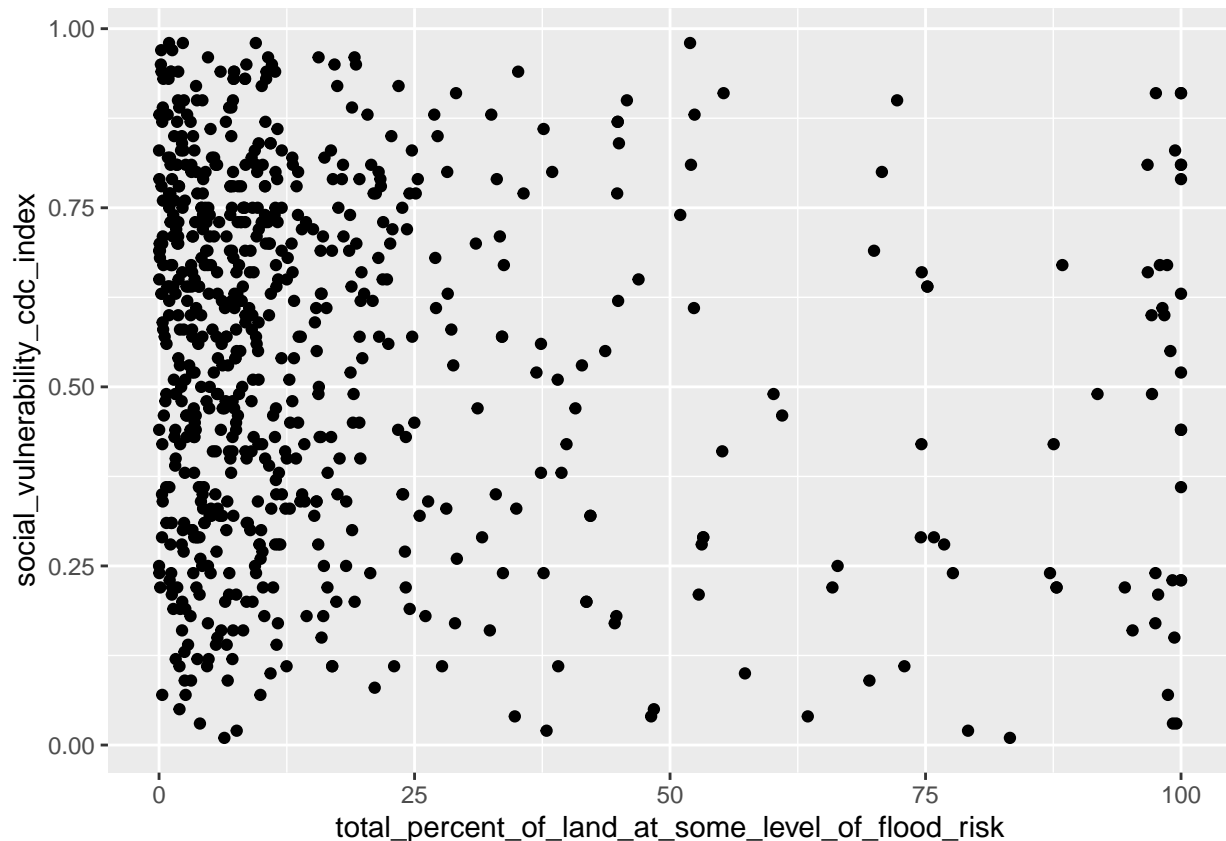
```
## Warning: Removed 93 rows containing non-finite values (`stat_bin()`).
```



```
#Scatterplot
```

```
ggplot(data, aes(x=total_percent_of_land_at_some_level_of_flood_risk, y=social_vulnerability_cdc_index)) +  
  geom_point()
```

```
## Warning: Removed 93 rows containing missing values (`geom_point()`).
```



Interpretation of missingness: There a large amount of missingnes in this data at 15% of the data missing. Variables which were missing over 70% of their data include: Place ID, County, City/Town, Encumbered (\$), Grant Award (4), Award Date, Funding Decision Year, Total Project Budget, Funding Request, StateAgency REquested for Funding, Application Date, Application Number, Project Title, Projcet Description, Applicant, Applicant Type, Project Type. The missing variables came from the grant data that was provided by the Golden Leaf Foundation.

Interpretation of Plots: The total percent of land at some level of flood risk is right skewed and also does not seem to be correlated with the CDC social vulnerability index. I chose to do some plots with this variable because I feel like it offers the most insight into the level of flooding in each community.

### RQ1: What communities have completed and submitted applications for funding?

- Within these applicants, what communities received funding?
- What are their characteristics?
- What type of funding did these communities receive?
- Who are key people to their application?
- What organizational infrastructure do they have to support the deployment of funding?

```
counties1_withdata <- data %>% na.omit()
```

### RQ2: How is funding distributed by grant administrators?

- Is there a correlation between communities funded and the wealth of the community and funding?

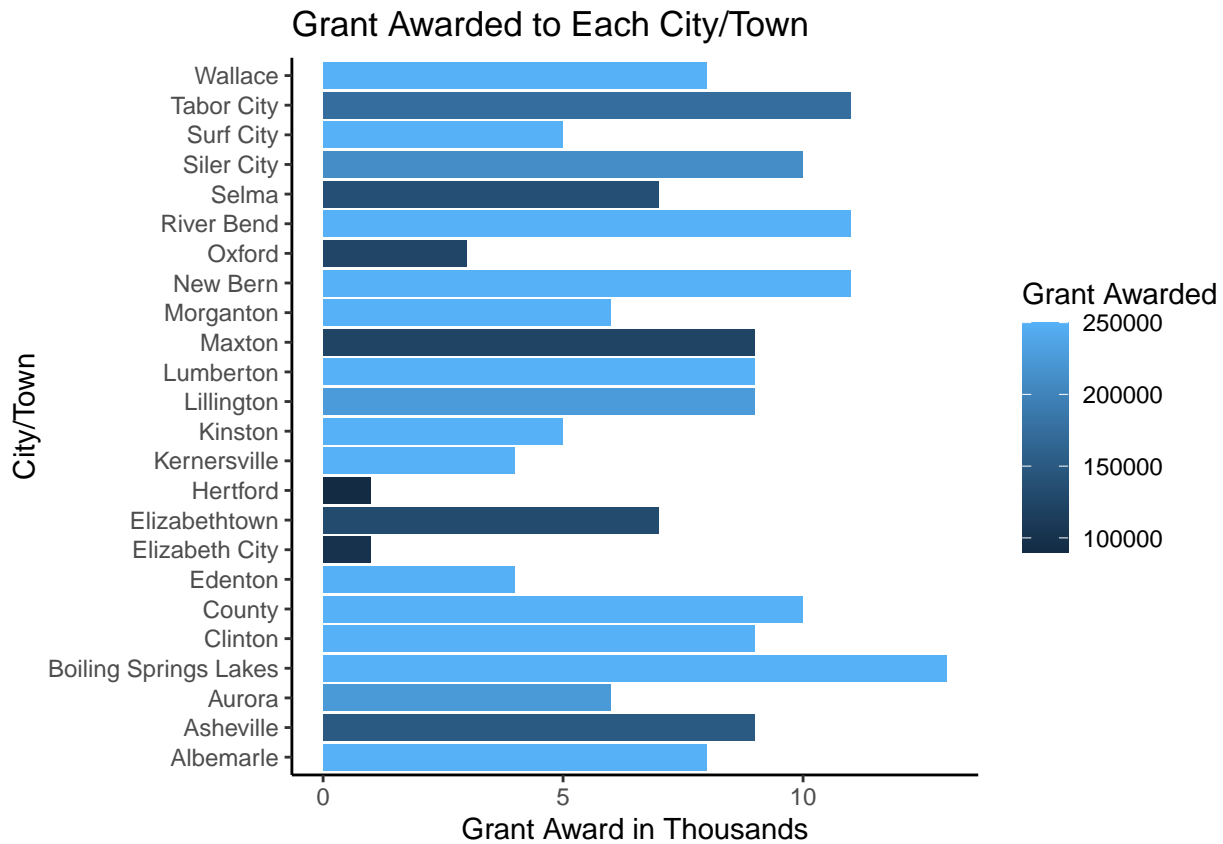
### RQ3: What trends exist in the data?

- What variables influence whether an organization receives funding?
- To what extent is there equitable distribution of funds?

- How does flood risk correlate with the amount of funds received if any?

*#Bar Plot visualizing which communities recieved grants and how much they received.*

```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(y=city_town, fill=grant_award)) + labs(y = "City/Town", x = "Grant Award in Thousands", ti
```

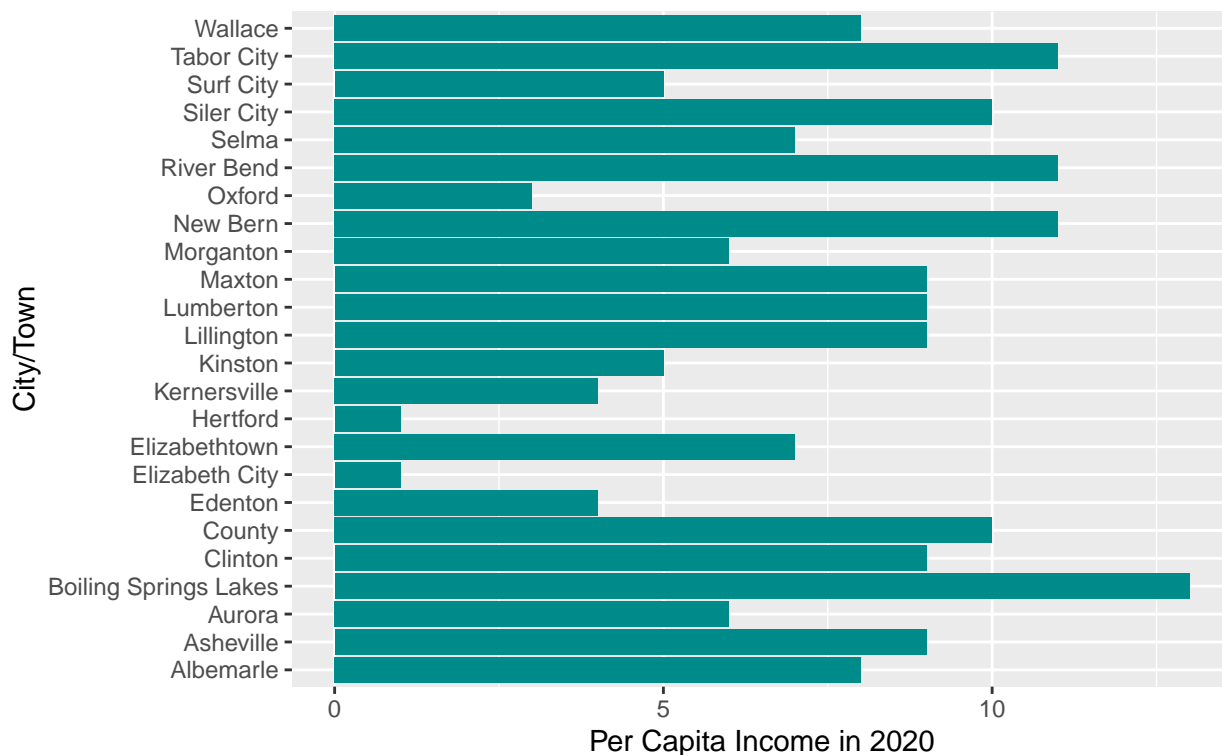


*#Bar Plot visualizing the Per Capita Income of communities who Recieved Grants*

```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(y=city_town, fill=per_capita_income_in_2020_inflation_adjusted_dollars)) + labs(y = "City/
```

## Per Capita Income of communities who Recieved Grants

### Income in 2020 Inflation Adjusted Dollars



*#Histograms visualizing the Per Capita income in 2020 of the communities which received funding*

```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(x=per_capita_income_in_2020_inflation_adjusted_dollars, fill=city_town), show.legend = FALSE)
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_density()`).
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Groups with fewer than two data points have been dropped.
```

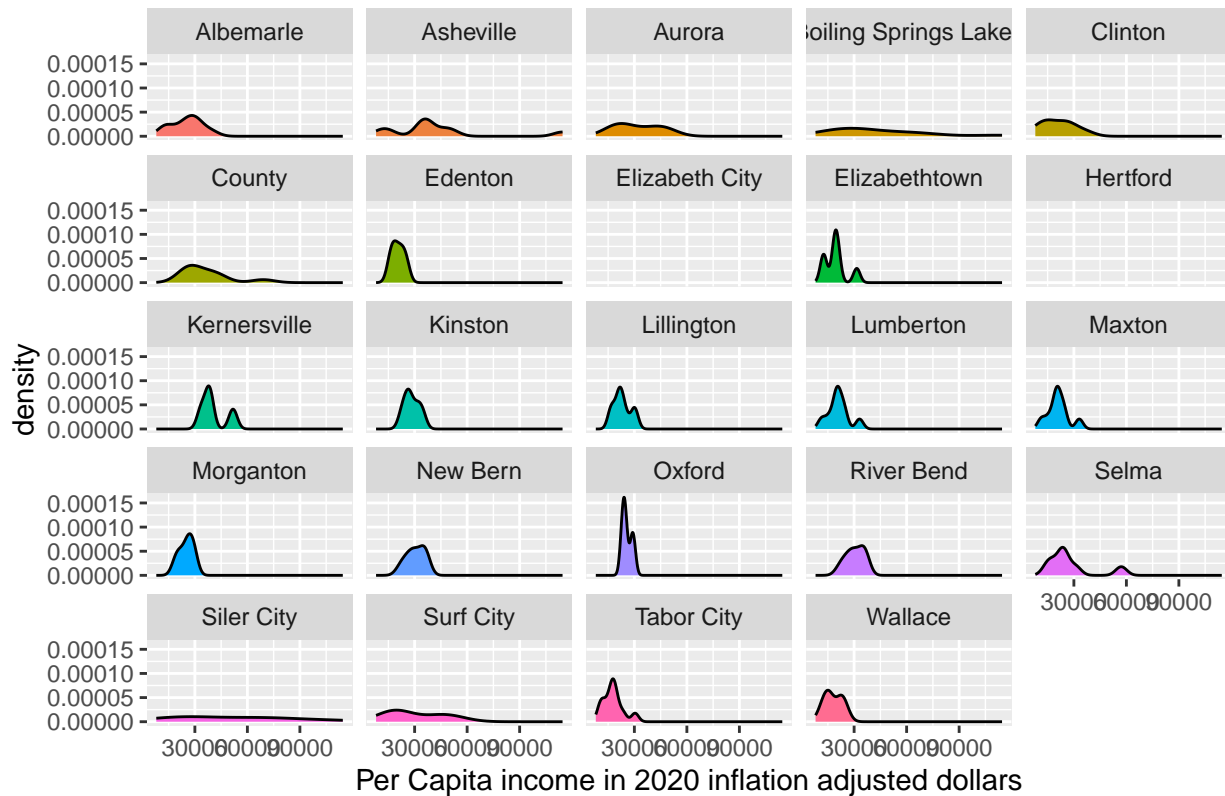
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
```

```
## -Inf
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
```

```
## -Inf
```

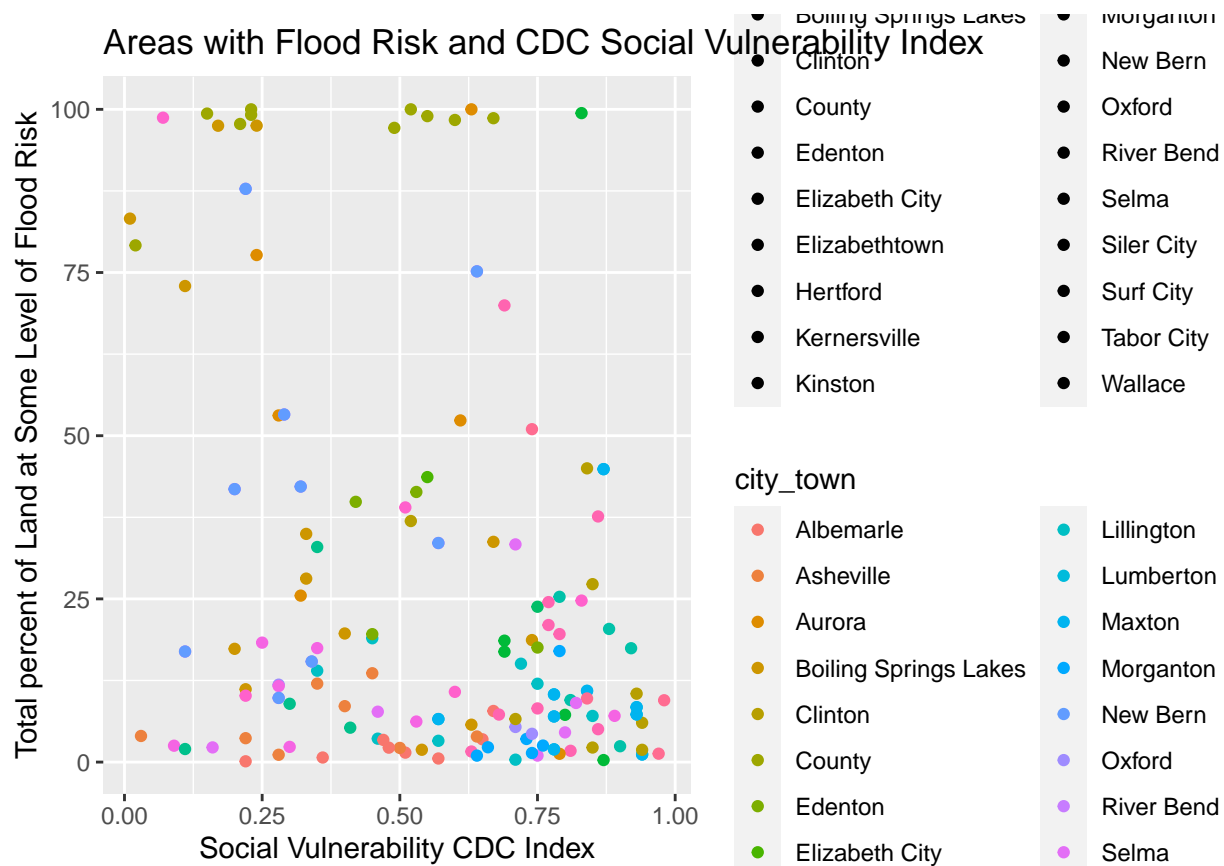
## Per Capita Income in 2020 of Communities Which Received a Grant



*#Scatterplot visualizing the relationship between social vulnerability index and percent of area flooded*

```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(x=social_vulnerability_cdc_index, y=total_percent_of_land_at_some_level_of_flood_risk, fill=city_town))
```

## Warning: Removed 18 rows containing missing values (`geom\_point()`).

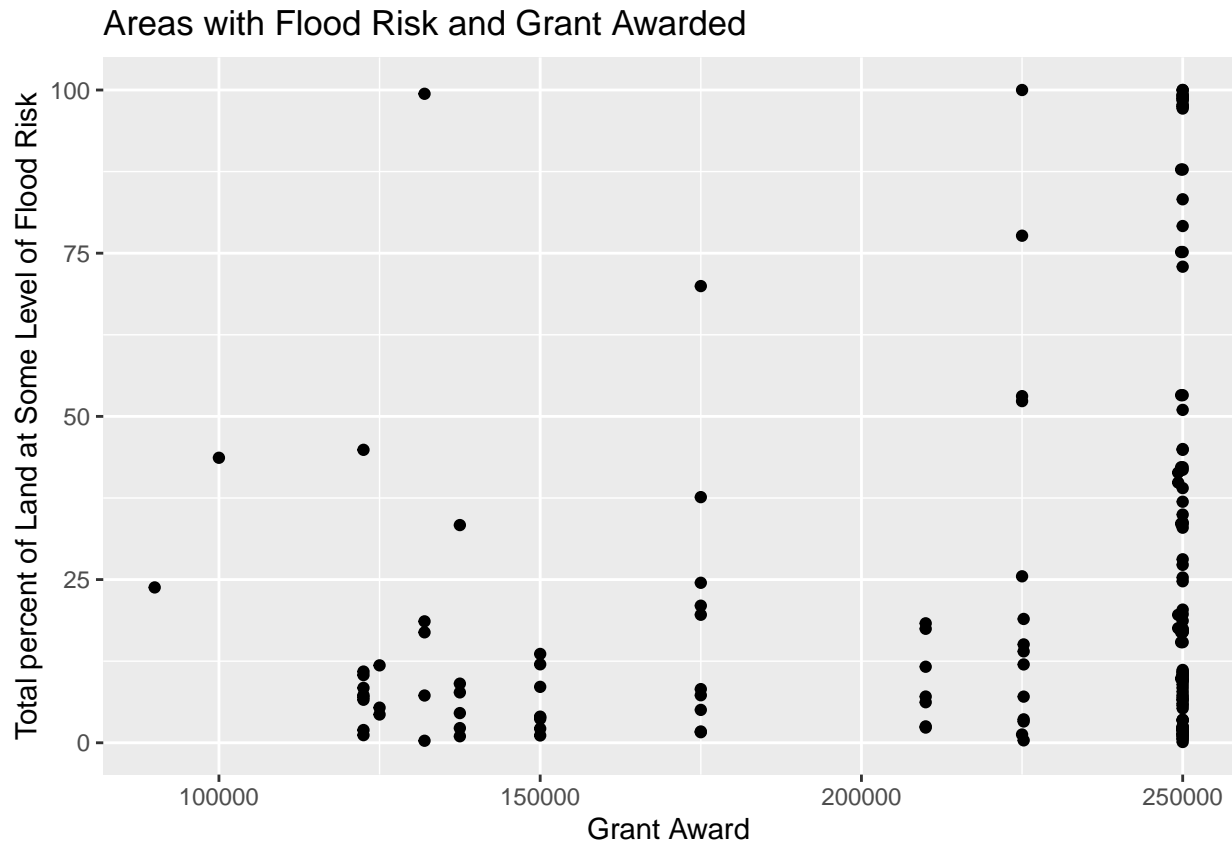


*#Scatterplot visualizing the relationship between grant awarded and percent of area flooded*

```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(x=grant_award, y=total_percent_of_land_at_some_level_of_flood_risk)) + labs(x = "Grant Awarded", y = "Total percent of Land at Some Level of Flood Risk")
```

## Warning: Removed 18 rows containing missing values (`geom\_point()`).





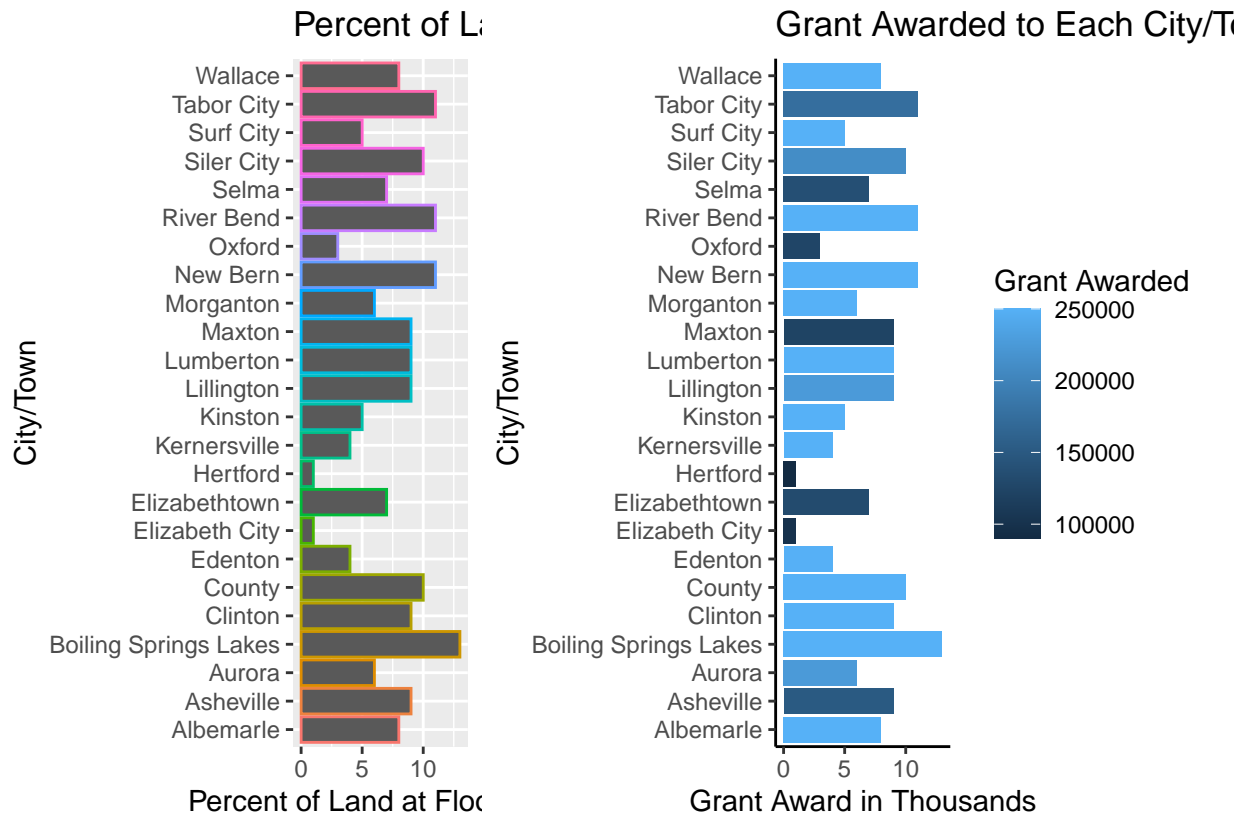
**RQ4: Is there a correlation between the size of the town and the amount of funds awarded?**

```
p1 <-
  data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(fill=total_percent_of_land_at_some_level_of_flood_risk, y=city_town, color=city_town)) + labs(
    y = "City/Town", x = "Grant Award in Thousands", title = "Areas with Flood Risk and Grant Awarded")

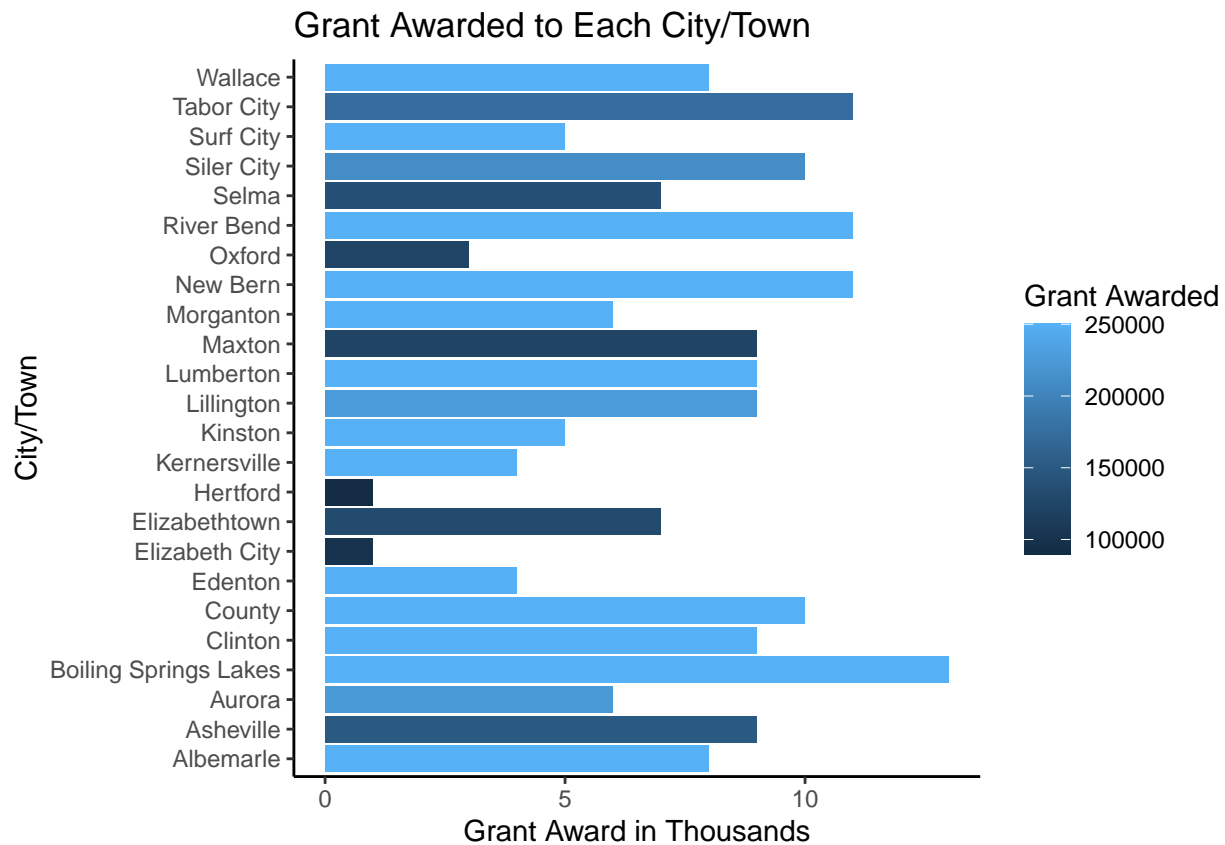
p2 <- data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(y=city_town, fill=grant_award)) + labs(y = "City/Town", x = "Grant Award in Thousands", title = "Areas with Flood Risk and Grant Awarded")

p1 + p2
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
## variable into a factor?
```



```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(y=city_town, fill=grant_award)) + labs(y = "City/Town", x = "Grant Award in Thousands", ti
```



```
data %>%
  filter(!is.na(city_town)) %>%
  filter(!is.na(grant_award)) %>%
  ggplot(aes(fill=total_percent_of_land_at_some_level_of_flood_risk, y=city_town, color=city_town)) +
```

## Warning: The following aesthetics were dropped during statistical transformation: fill

## i This can happen when ggplot fails to infer the correct grouping structure in

## the data.

## i Did you forget to specify a `group` aesthetic or to convert a numerical

## variable into a factor?

# Percent of Land at Flood Risk in each City/Town

## Communtities Which Recieved Funding

