

Data Profiling Report

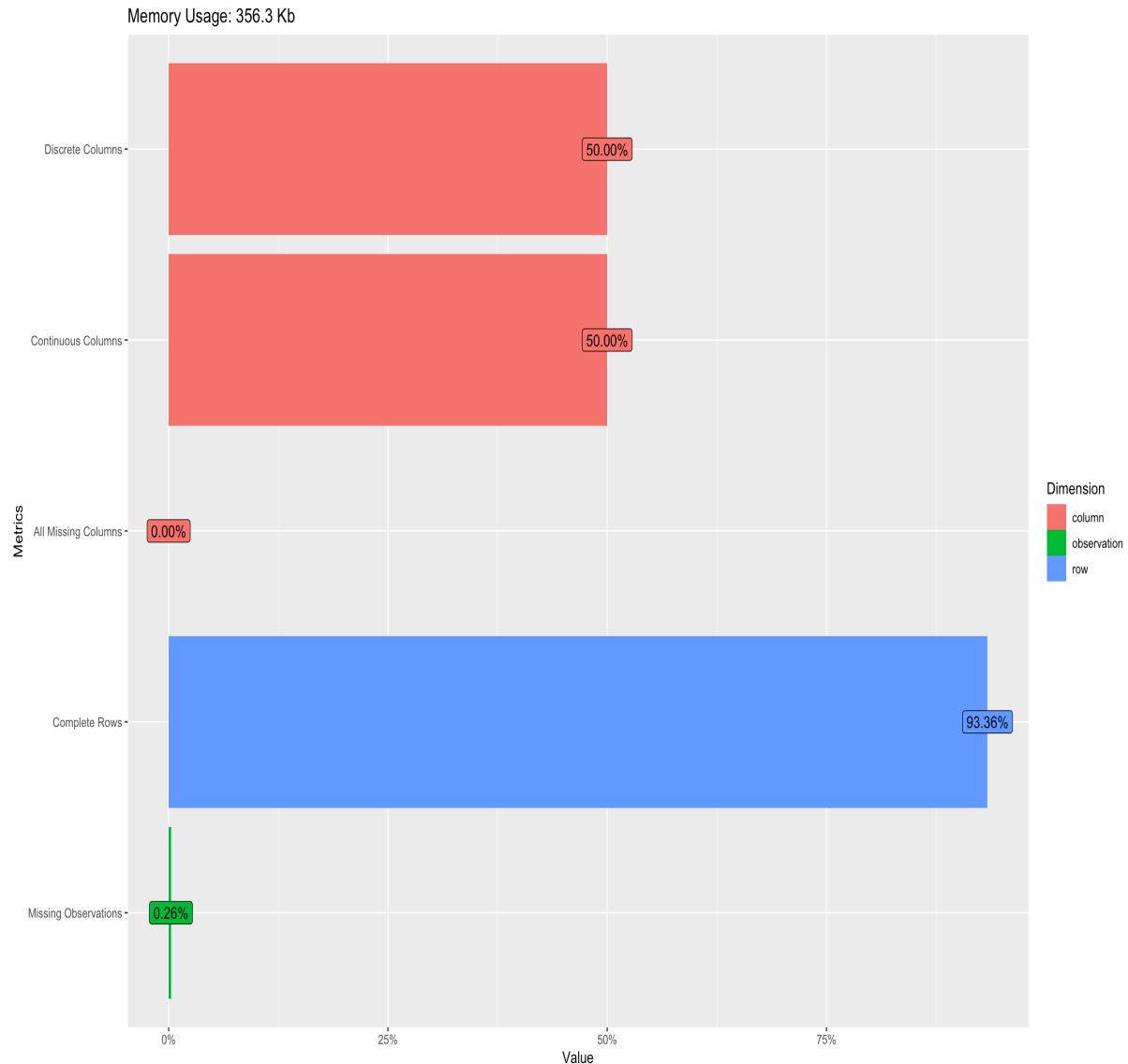
- [Basic Statistics](#)
 - [Raw Counts](#)
 - [Percentages](#)
- [Data Structure](#)
- [Missing Data Profile](#)
- [Univariate Distribution](#)
 - [Histogram](#)
 - [Bar Chart \(with frequency\)](#)
 - [QQ Plot](#)
- [Correlation Analysis](#)
- [Principal Component Analysis](#)

Basic Statistics

Raw Counts

Name	Value
Rows	467
Columns	34
Discrete columns	17
Continuous columns	17
All missing columns	0
Missing observations	41
Complete Rows	436
Total observations	15,878
Memory allocation	356.3 Kb

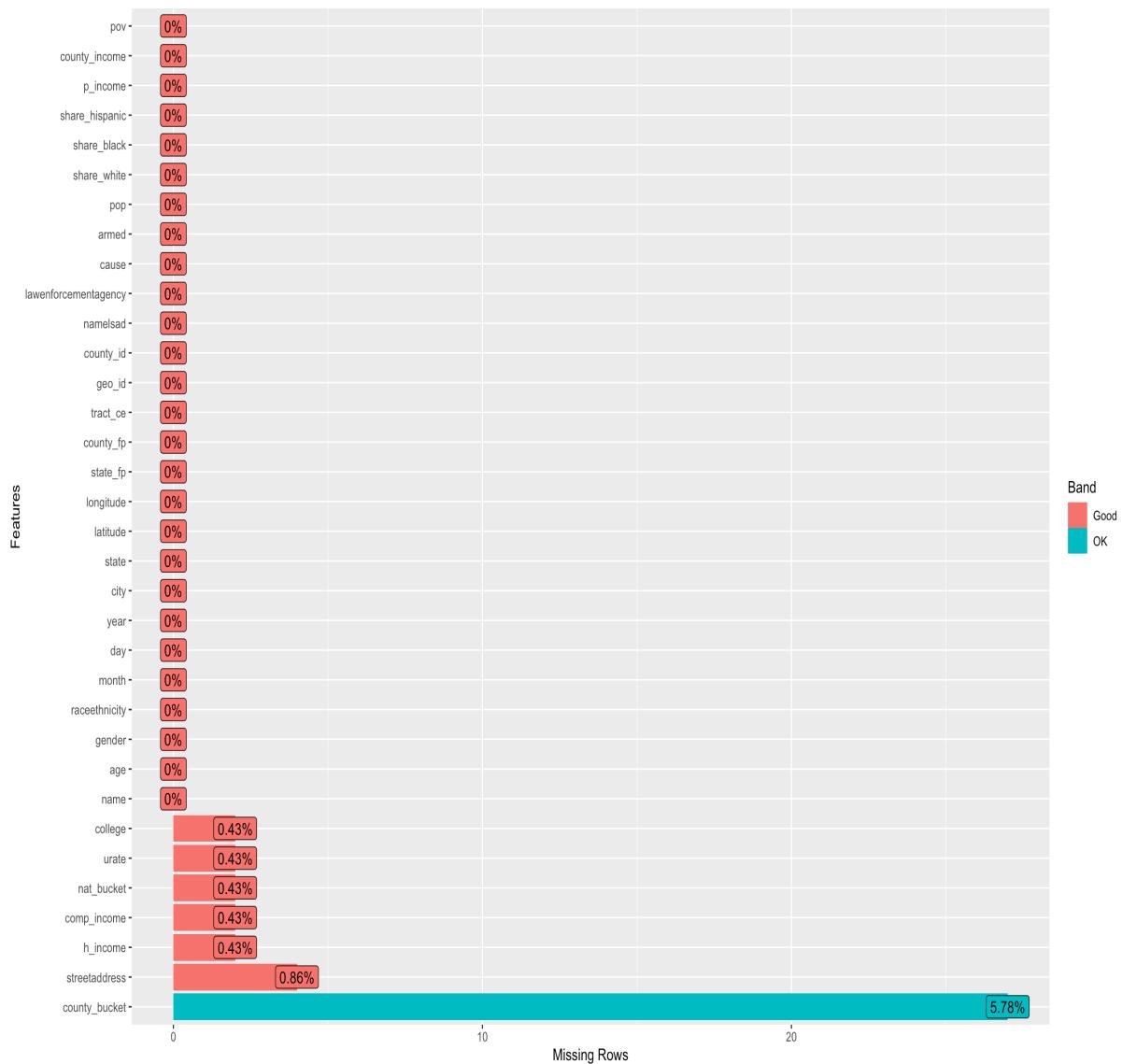
Percentages



Data Structure

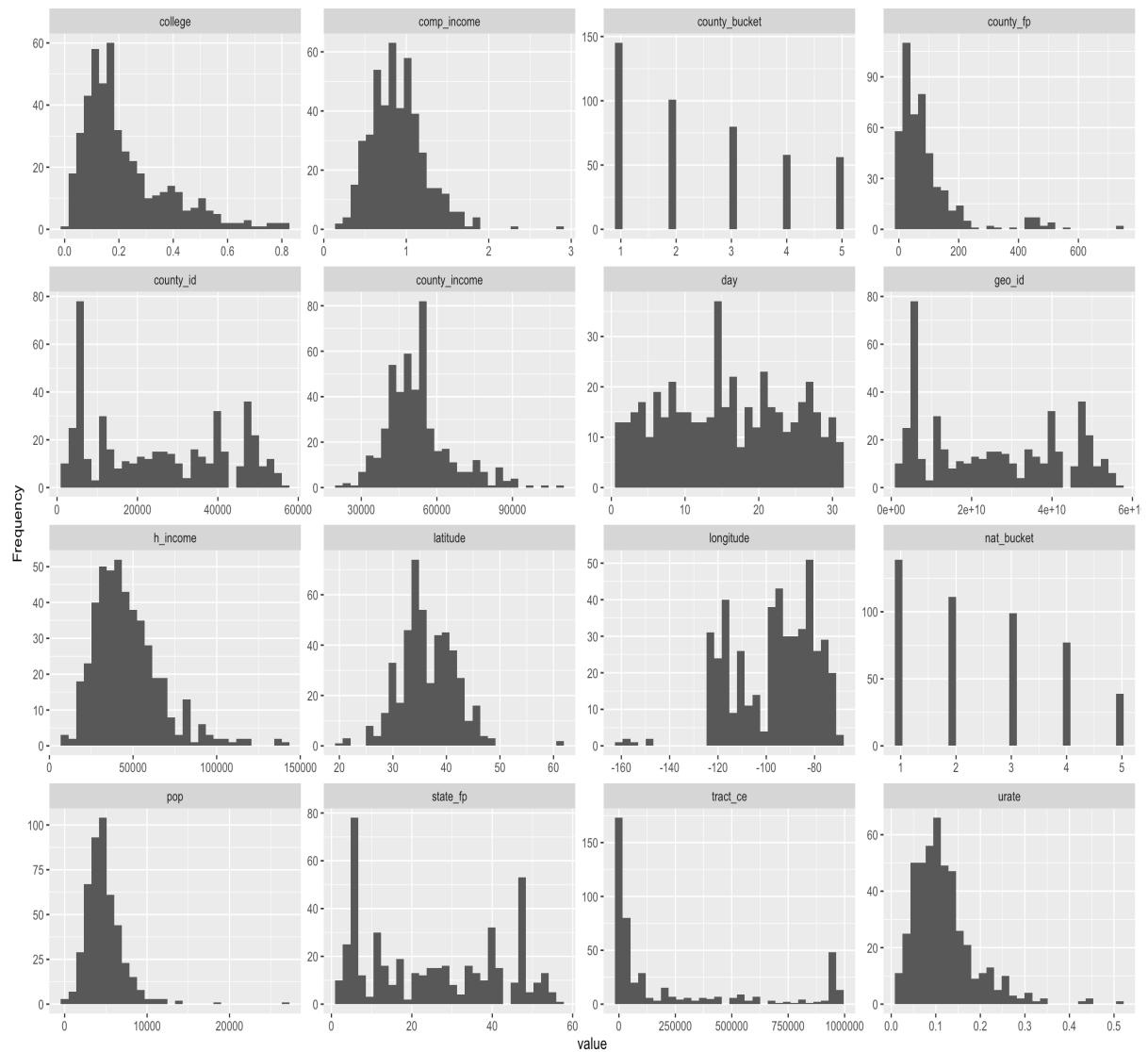
```
root (Classes 'data.table' and 'data.frame': 467 obs. of 34 variables:)  
  - name (chr) age (chr) gender (chr) raceethnicity (chr) month (chr) day (int) year (int) streetaddress (chr)  
  - city (chr) state (chr) latitude (num) longitude (num) state_fp (int) county_fp (int) tract_ce (int) geo_id (num) county_id (int) name (lsad (chr)) law enforcement agency (chr) cause (chr) armed (chr) pop (int) share_white (chr) share_black (chr) share_hispanic (chr) p_income (chr) h_income (int) county_income (int) comp_income (num) county_bucket (int) nat_bucket (int) pov (chr) urate (num) college (num)
```

Missing Data Profile



Univariate Distribution

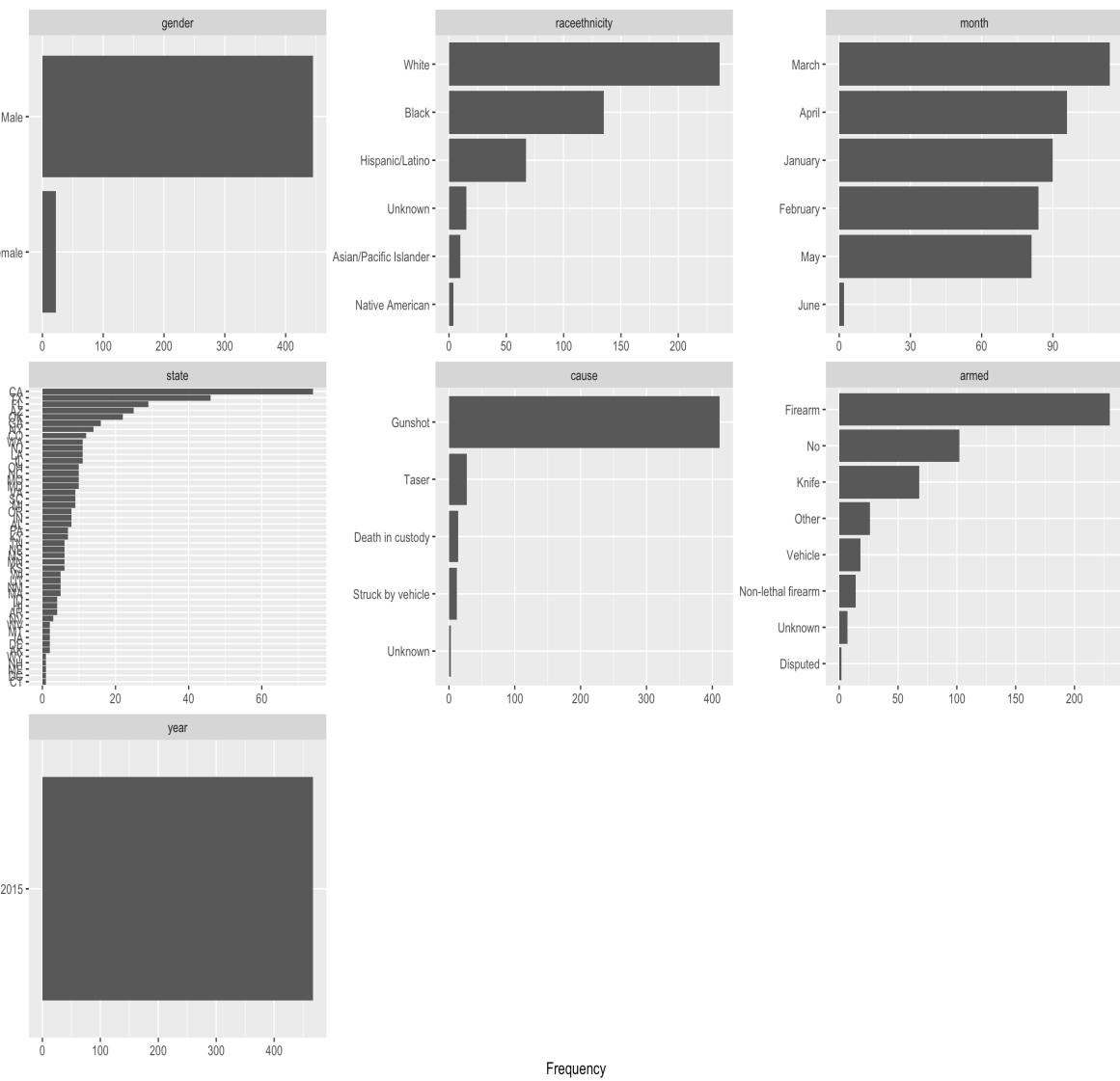
Histogram



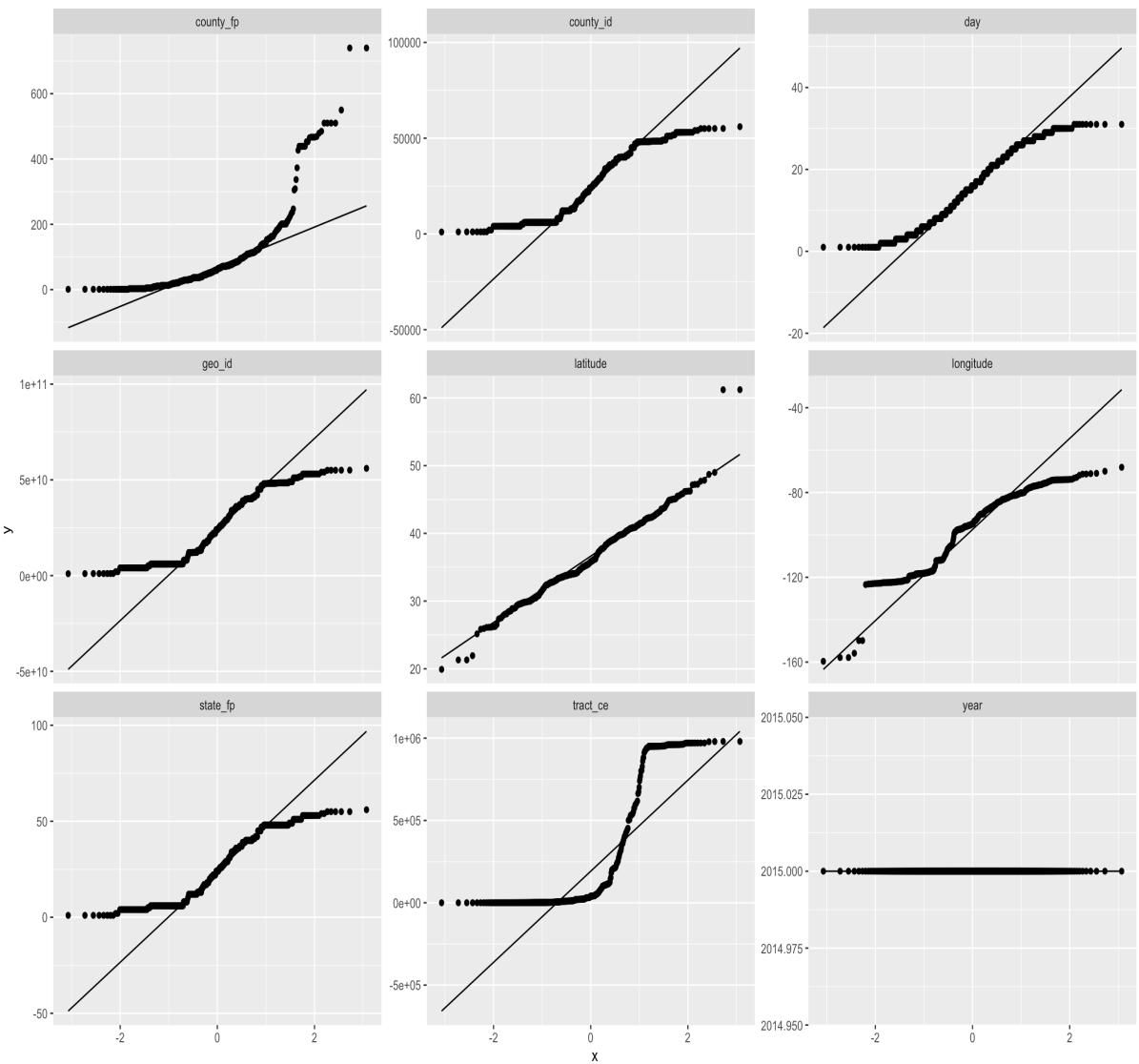
Bar Chart (with frequency)

```
## 11 columns ignored with more than 50 categories.
## name: 465 categories
## age: 61 categories
## streetaddress: 460 categories
## city: 364 categories
## namelsad: 389 categories
## lawenforcementagency: 377 categories
## share_white: 363 categories
## share_black: 246 categories
## share_hispanic: 293 categories
## p_income: 452 categories
```

```
## pov: 281 categories
```

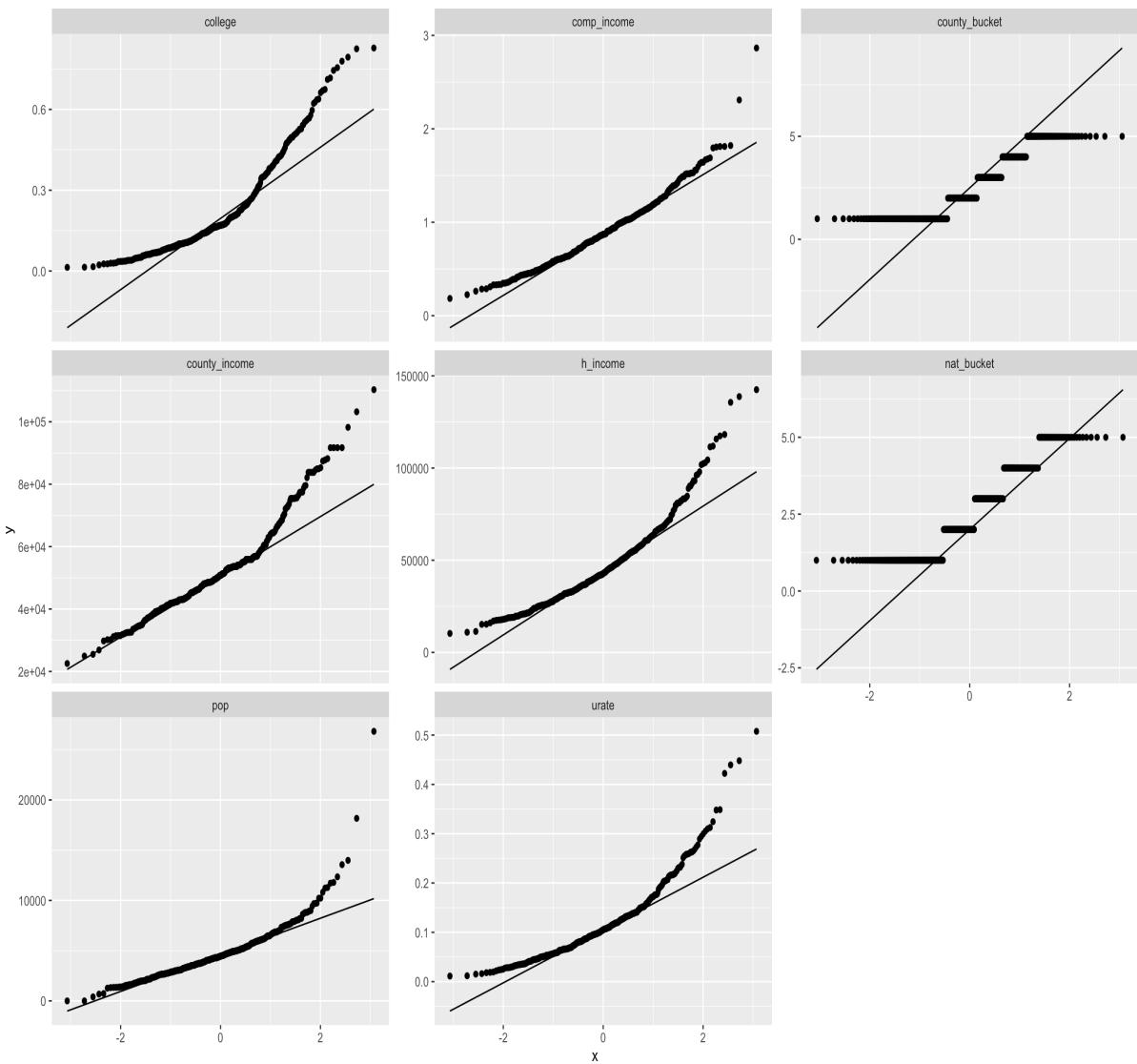


QQ Plot



Page 1

```
#> #> Warning: Removed 37 rows containing non-finite values (stat_qq).
#> #> Warning: Removed 37 rows containing non-finite values (stat_qq_line).
```



Page 2

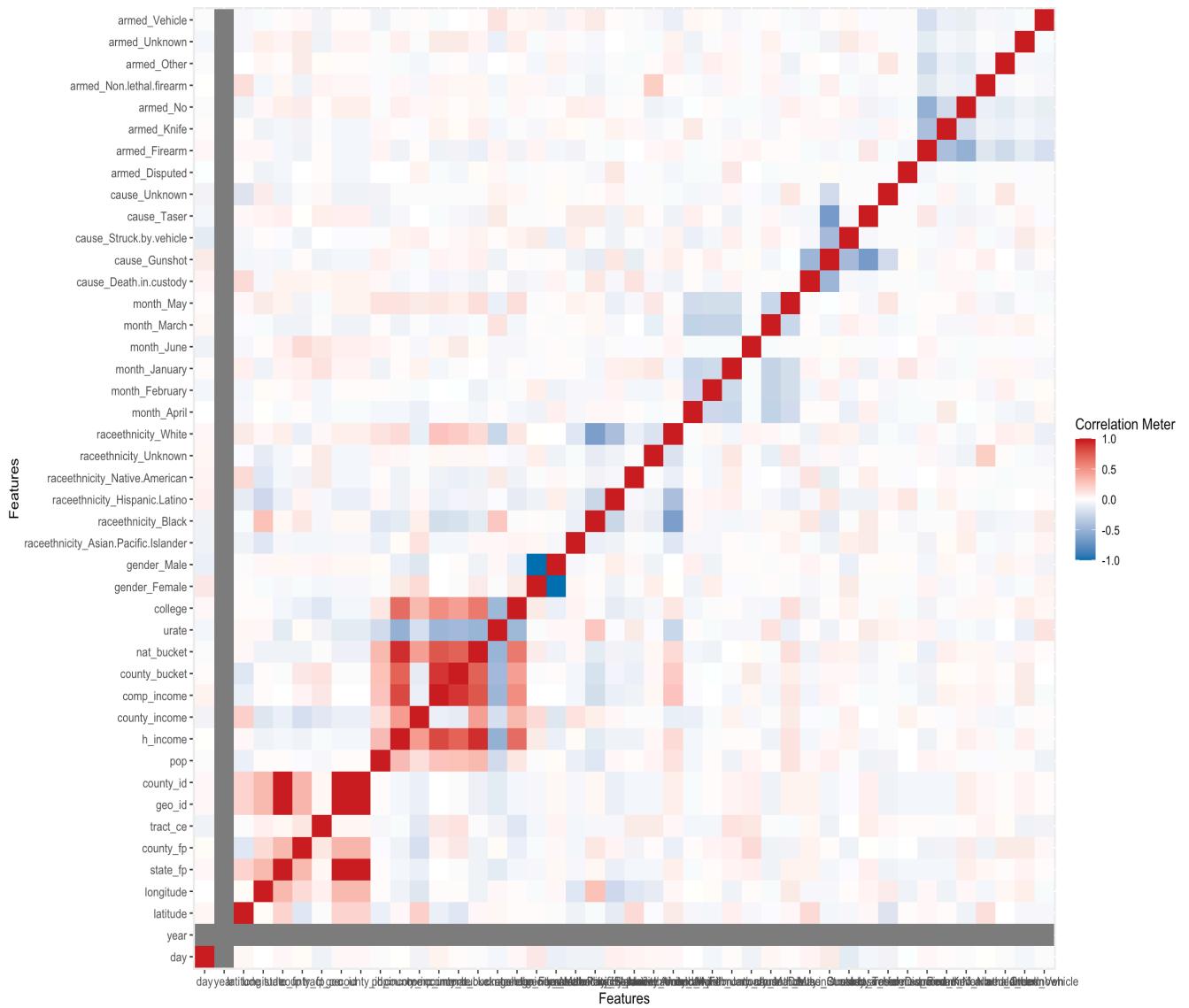
Correlation Analysis

```
## 12 features with more than 20 categories ignored!
## name: 434 categories
## age: 60 categories
## streetaddress: 433 categories
## city: 336 categories
## state: 47 categories
## namelsad: 371 categories
## lawenforcementagency: 349 categories
## share_white: 347 categories
## share_black: 237 categories
## share_hispanic: 285 categories
```

```

## p_income: 422 categories
## pov: 267 categories
## Warning in cor(x = structure(list(day = c(23L, 2L, 14L, 11L, 19L, 7L, 27L, : the standard deviation is zero

```



Principal Component Analysis

```

## 11 features with more than 50 categories ignored!
## name: 434 categories
## age: 60 categories
## streetaddress: 433 categories
## city: 336 categories
## namelsad: 371 categories
## lawenforcementagency: 349 categories

```

```
## share_white: 347 categories
## share_black: 237 categories
## share_hispanic: 285 categories
## p_income: 422 categories
## pov: 267 categories
## Warning in plot_prcp(data = structure(list(name = c("A'donte Washington",
## : The following features are dropped due to zero variance:
## * year
```

% Variance Explained By Principal Components
 (Note: Labels indicate cumulative % explained variance)

