R Notebook

Isabel Metzger Homework 4_4 Data Wrangling

```
library(readr)
stategrid <- read_delim("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machinelearning_IM_MG/
    "\t", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   state = col_character(),
##   x = col_integer(),
##   y = col_integer()
## )
```

```
ACS_14_5YR_DP02_ <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machinelearning_I
    skip = 1)
```

```
## Multiple files in zip: reading 'ACS_14_5YR_DP02_with_ann.csv'
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Id = col_character(),
##   Id2 = col_character(),
##   Geography = col_character(),
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households` = col_character(),
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families)` = col_double(),
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families)` = c
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - With own children
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Wi
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Married-couple fam
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Ma
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Married-couple fam
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Ma
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Male householder,
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Ma
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Male householder,
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Ma
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Female householde
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Fe
##   `Percent; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Female householde
##   `Percent Margin of Error; HOUSEHOLDS BY TYPE - Total households - Family households (families) - Fe
##   # ... with 281 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
## Warning: package 'purrr' was built under R version 3.4.2
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
trim <- function (x) gsub("^\\s+|\\s+$", "", x)   # function to trim spaces of columns
multi.fun <- function(x) {cbind(freq = table(x), percentage = prop.table(table(x))*100)}
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.2
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(readr)
stategrid <- read_delim("state-grid-coordinates.tsv",
    "\t", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   state = col_character(),
##   x = col_integer(),
##   y = col_integer()
## )
```

```
# loading various df
#gov datasets
opioids_name_list <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machinelearning
```

```
## Parsed with column specification:
## cols(
##   `Drug Name` = col_character(),
##   `Generic Name` = col_character()
## )
```

```
OD_Multiple_Cause_of_Death_1999_2014_v1_1 <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_projec
```

```
## Parsed with column specification:
## cols(
##   State = col_character(),
##   Year = col_integer(),
##   Deaths = col_character(),
##   Population = col_integer(),
##   `Crude Rate` = col_character(),
##   `Crude Rate Lower 95% Confidence Interval` = col_character(),
##   `Crude Rate Upper 95% Confidence Interval` = col_character(),
```

```
##   `Prescriptions Dispensed by US Retailers in that year (millions)` = col_integer()
## )
```

```
Opioid_analgesic_prescriptions_dispensed_from_US_retail_pharmacies_Q4_2009_Q2_2015 <- read_csv("~/R_STU
```

```
## Parsed with column specification:
## cols(
##   Quarter = col_character(),
##   Hydrocodone = col_integer(),
##   Oxycodone = col_integer(),
##   Tramadol = col_integer(),
##   Morphine = col_integer(),
##   Fentanyl = col_integer(),
##   `H+O+T+M+F` = col_integer(),
##   `All Opioid Analgesics` = col_integer(),
##   `ER/LA Opioid Analgesics` = col_integer(),
##   `Yearly totals (All Opioid Analgesics)` = col_character(),
##   `Yearly totals (ER/LA Opioid Analgesics)` = col_character(),
##   `Yearly totals (H+O+T+M+F)` = col_character(),
##   `Yearly totals (H+O)` = col_character()
## )
```

```
ACS_14_5YR_B07012_PovertyState <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_ma
    skip = 1)
```

```
## Multiple files in zip: reading 'ACS_14_5YR_B07012_with_ann.csv'
```

```
## Parsed with column specification:
## cols(
##   Id = col_character(),
##   Id2 = col_character(),
##   Geography = col_character(),
##   `Estimate; Total:` = col_integer(),
##   `Estimate; Total: - Below 100 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - 100 to 149 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - At or above 150 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - Same house 1 year ago:` = col_integer(),
##   `Estimate; Total: - Same house 1 year ago: - At or above 150 percent of the poverty level` = col_i
##   `Estimate; Total: - Moved within same county:` = col_integer(),
##   `Estimate; Total: - Moved within same county: - At or above 150 percent of the poverty level` = co
##   `Estimate; Total: - Moved from different county within same state:` = col_integer(),
##   `Estimate; Total: - Moved from different county within same state: - At or above 150 percent of th
##   `Estimate; Total: - Moved from different state:` = col_integer(),
##   `Estimate; Total: - Moved from different state: - 100 to 149 percent of the poverty level` = col_i
##   `Estimate; Total: - Moved from abroad:` = col_integer(),
##   `Estimate; Total: - Moved from abroad: - 100 to 149 percent of the poverty level` = col_integer()
## )
```

```
ACS_13_5YR_B07012_Poverty <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machine
    skip = 1)
```

```
## Multiple files in zip: reading 'ACS_13_5YR_B07012_with_ann.csv'
```

```
## Parsed with column specification:
## cols(
##   Id = col_character(),
##   Id2 = col_character(),
```

```
##   Geography = col_character(),
##   `Estimate; Total:` = col_integer(),
##   `Estimate; Total: - Below 100 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - 100 to 149 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - At or above 150 percent of the poverty level` = col_integer(),
##   `Estimate; Total: - Same house 1 year ago:` = col_integer(),
##   `Estimate; Total: - Same house 1 year ago: - At or above 150 percent of the poverty level` = col_i
##   `Estimate; Total: - Moved within same county:` = col_integer(),
##   `Estimate; Total: - Moved within same county: - 100 to 149 percent of the poverty level` = col_inte
##   `Estimate; Total: - Moved from different county within same state:` = col_integer(),
##   `Estimate; Total: - Moved from different county within same state: - 100 to 149 percent of the pove
##   `Estimate; Total: - Moved from different state:` = col_integer(),
##   `Estimate; Total: - Moved from different state: - 100 to 149 percent of the poverty level` = col_i
##   `Estimate; Total: - Moved from abroad:` = col_integer(),
##   `Estimate; Total: - Moved from abroad: - At or above 150 percent of the poverty level` = col_intege
## )
```

```
OD <- read.csv('overdoses_2014.csv')
```

```
library(readxl)
Part_D_Opioid_Prescribing_Change_Geographic_2013_2014 <- read_excel("~/R_STUDIO_FALL_2017_PDA/hw2-izzyka
    skip = 2)
```

```
str(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    51 obs. of  12 variables:
##  $ State Name                                          : chr  "Alabama" "Alaska" "Arizona" "Arkansa
##  $ State Abbreviation                                  : chr  "AL" "AK" "AZ" "AR" ...
##  $ State FIPS                                          : chr  "01" "02" "04" "05" ...
##  $ Part D Prescribers 2013                             : num  12820 2275 20542 7909 109571 ...
##  $ Opioid Claims 2013                                  : num  2260284 86517 1545138 1128356 722878
##  $ Overall Claims 2013                                 : num  2.92e+07 1.28e+06 2.21e+07 1.68e+07
##  $ Opioid Prescribing Rate 2013                        : num  7.75 6.75 6.98 6.73 5.48 ...
##  $ Part D Prescribers 2014                             : num  13100 2293 21253 8010 111395 ...
##  $ Opioid Claims 2014                                  : num  2267090 93606 1639782 1147588 734152
##  $ Overall Claims 2014                                 : num  2.89e+07 1.38e+06 2.35e+07 1.73e+07
##  $ Opioid Prescribing Rate 2014                        : num  7.86 6.76 6.99 6.63 5.46 ...
##  $ Percentage Point Difference in Opioid Prescribing Rate: num  0.10639 0.00767 0.00797 -0.10019 -0.0
```

```
library(tidyr)
library(dplyr)
```

```
library(tidyverse)
OpioidPrescribingTibble13.14 <- as_tibble(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014)
```

```
setnames(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014, "State Name", "State")
```

```
OpioidPrescribingTibble13.14 %>% .[["Opioid Claims 2013"]]
```

```
##  [1] 2260284   86517 1545138 1128356 7228782 1084129  671688  233078
##  [9]   81482 5246555 2703771  176305  435716 2292480 2125966  805242
## [17]  860248 1763873 1559052  425795  939000 1331244 3328161 1082073
## [25] 1101786 2039124  268035  454065  681809  293762 1447669  442975
## [33] 2804472 3204555  164414 3447253 1298037 1206329 3530705  254915
## [41] 1439682  209426 2875441 5182878  522285  144746 1711665 1575510
## [49]  808720 1444615   95875
```

```
x <- type_convert(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014)  # table_df
```

```
## Parsed with column specification:
## cols(
##   State = col_character(),
##   `State Abbreviation` = col_character(),
##   `State FIPS` = col_character()
## )
```

```
library(readr)
PEP_2016_PEPANNRES <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machinelearnin
    skip = 1)
```

```
## Multiple files in zip: reading 'PEP_2016_PEPANNRES_with_ann.csv'
```

```
## Parsed with column specification:
## cols(
##   Id = col_character(),
##   Id2 = col_character(),
##   Geography = col_character(),
##   `April 1, 2010 - Census` = col_integer(),
##   `April 1, 2010 - Estimates Base` = col_integer(),
##   `Population Estimate (as of July 1) - 2010` = col_integer(),
##   `Population Estimate (as of July 1) - 2011` = col_integer(),
##   `Population Estimate (as of July 1) - 2012` = col_integer(),
##   `Population Estimate (as of July 1) - 2013` = col_integer(),
##   `Population Estimate (as of July 1) - 2014` = col_integer(),
##   `Population Estimate (as of July 1) - 2015` = col_integer(),
##   `Population Estimate (as of July 1) - 2016` = col_integer()
## )
```

```
head(PEP_2016_PEPANNRES, 3)
```

```
## # A tibble: 3 x 12
##           Id   Id2        Geography `April 1, 2010 - Census`
##        <chr> <chr>            <chr>                    <int>
## 1  0100000US  <NA>    United States                308745538
## 2 0200000US1     1 Northeast Region                 55317240
## 3 0200000US2     2   Midwest Region                 66927001
## # ... with 8 more variables: `April 1, 2010 - Estimates Base` <int>,
## #   `Population Estimate (as of July 1) - 2010` <int>, `Population
## #   Estimate (as of July 1) - 2011` <int>, `Population Estimate (as of
## #   July 1) - 2012` <int>, `Population Estimate (as of July 1) -
## #   2013` <int>, `Population Estimate (as of July 1) - 2014` <int>,
## #   `Population Estimate (as of July 1) - 2015` <int>, `Population
## #   Estimate (as of July 1) - 2016` <int>
```

```
fullUS_NE_MW_SOUTH_WEST <- PEP_2016_PEPANNRES[PEP_2016_PEPANNRES$Id == '0100000US' | PEP_2016_PEPANNRES$
```

```
fullUS_NE_MW_SOUTH_WEST
```

```
## # A tibble: 5 x 12
##           Id   Id2        Geography `April 1, 2010 - Census`
##        <chr> <chr>            <chr>                    <int>
## 1  0100000US  <NA>    United States                308745538
## 2 0200000US1     1 Northeast Region                 55317240
## 3 0200000US2     2   Midwest Region                 66927001
```

```
## 4 0200000US3     3     South Region                114555744
## 5 0200000US4     4      West Region                 71945553
## # ... with 8 more variables: `April 1, 2010 - Estimates Base` <int>,
## #   `Population Estimate (as of July 1) - 2010` <int>, `Population
## #   Estimate (as of July 1) - 2011` <int>, `Population Estimate (as of
## #   July 1) - 2012` <int>, `Population Estimate (as of July 1) -
## #   2013` <int>, `Population Estimate (as of July 1) - 2014` <int>,
## #   `Population Estimate (as of July 1) - 2015` <int>, `Population
## #   Estimate (as of July 1) - 2016` <int>
```

```
colnames(PEP_2016_PEPANNRES)
```

```
##  [1] "Id"
##  [2] "Id2"
##  [3] "Geography"
##  [4] "April 1, 2010 - Census"
##  [5] "April 1, 2010 - Estimates Base"
##  [6] "Population Estimate (as of July 1) - 2010"
##  [7] "Population Estimate (as of July 1) - 2011"
##  [8] "Population Estimate (as of July 1) - 2012"
##  [9] "Population Estimate (as of July 1) - 2013"
## [10] "Population Estimate (as of July 1) - 2014"
## [11] "Population Estimate (as of July 1) - 2015"
## [12] "Population Estimate (as of July 1) - 2016"
```

```
select.group <- c("Id", "Geography","Population Estimate (as of July 1) - 2013","Population Estimate (as
CENSUS_13.14 <- select(PEP_2016_PEPANNRES, select.group)
```

```
head(CENSUS_13.14, 3)
```

```
## # A tibble: 3 x 4
##           Id        Geography `Population Estimate (as of July 1) - 2013`
##        <chr>            <chr>                                       <int>
## 1  0100000US    United States                                   316204908
## 2 0200000US1 Northeast Region                                    55988771
## 3 0200000US2   Midwest Region                                    67543948
## # ... with 1 more variables: `Population Estimate (as of July 1) -
## #   2014` <int>
```

```
prescriber_info <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/prescriber-info.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Gender = col_character(),
##   State = col_character(),
##   Credentials = col_character(),
##   Specialty = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
Opioid.PrescribersDF <- prescriber_info[prescriber_info$Opioid.Prescriber == 1,]
```

```
head(Opioid.PrescribersDF)
```

```
## # A tibble: 6 x 256
##          NPI Gender State Credentials           Specialty ABILIFY
```

```
##        <int>  <chr> <chr>       <chr>               <chr>   <int>
## 1 1710982582      M    TX         DDS             Dentist       0
## 2 1245278100      F    AL          MD     General Surgery       0
## 3 1669567541      M    AZ          MD   Internal Medicine       0
## 4 1679650949      M    NV        M.D. Hematology/Oncology       0
## 5 1548580897      M    PA          DO     General Surgery       0
## 6 1437192002      M    NH          MD     Family Practice       0
## # ... with 250 more variables: ACETAMINOPHEN.CODEINE <int>,
## #   ACYCLOVIR <int>, ADVAIR.DISKUS <int>, AGGRENOX <int>,
## #   ALENDRONATE.SODIUM <int>, ALLOPURINOL <int>, ALPRAZOLAM <int>,
## #   AMIODARONE.HCL <int>, AMITRIPTYLINE.HCL <int>,
## #   AMLODIPINE.BESYLATE <int>, AMLODIPINE.BESYLATE.BENAZEPRIL <int>,
## #   AMOXICILLIN <int>, AMOX.TR.POTASSIUM.CLAVULANATE <int>,
## #   AMPHETAMINE.SALT.COMBO <int>, ATENOLOL <int>,
## #   ATORVASTATIN.CALCIUM <int>, AVODART <int>, AZITHROMYCIN <int>,
## #   BACLOFEN <int>, BD.ULTRA.FINE.PEN.NEEDLE <int>, BENAZEPRIL.HCL <int>,
## #   BENICAR <int>, BENICAR.HCT <int>, BENZTROPINE.MESYLATE <int>,
## #   BISOPROLOL.HYDROCHLOROTHIAZIDE <int>, BRIMONIDINE.TARTRATE <int>,
## #   BUMETANIDE <int>, BUPROPION.HCL.SR <int>, BUPROPION.XL <int>,
## #   BUSPIRONE.HCL <int>, BYSTOLIC <int>, CARBAMAZEPINE <int>,
## #   CARBIDOPA.LEVODOPA <int>, CARISOPRODOL <int>, CARTIA.XT <int>,
## #   CARVEDILOL <int>, CEFUROXIME <int>, CELEBREX <int>, CEPHALEXIN <int>,
## #   CHLORHEXIDINE.GLUCONATE <int>, CHLORTHALIDONE <int>, CILOSTAZOL <int>,
## #   CIPROFLOXACIN.HCL <int>, CITALOPRAM.HBR <int>, CLINDAMYCIN.HCL <int>,
## #   CLOBETASOL.PROPIONATE <int>, CLONAZEPAM <int>, CLONIDINE.HCL <int>,
## #   CLOPIDOGREL <int>, CLOTRIMAZOLE.BETAMETHASONE <int>, COLCRYS <int>,
## #   COMBIVENT.RESPIMAT <int>, CRESTOR <int>, CYCLOBENZAPRINE.HCL <int>,
## #   DEXILANT <int>, DIAZEPAM <int>, DICLOFENAC.SODIUM <int>,
## #   DICYCLOMINE.HCL <int>, DIGOX <int>, DIGOXIN <int>,
## #   DILTIAZEM.24HR.CD <int>, DILTIAZEM.24HR.ER <int>, DILTIAZEM.ER <int>,
## #   DILTIAZEM.HCL <int>, DIOVAN <int>, DIPHENOXYLATE.ATROPINE <int>,
## #   DIVALPROEX.SODIUM <int>, DIVALPROEX.SODIUM.ER <int>,
## #   DONEPEZIL.HCL <int>, DORZOLAMIDE.TIMOLOL <int>,
## #   DOXAZOSIN.MESYLATE <int>, DOXEPIN.HCL <int>,
## #   DOXYCYCLINE.HYCLATE <int>, DULOXETINE.HCL <int>,
## #   ENALAPRIL.MALEATE <int>, ESCITALOPRAM.OXALATE <int>, ESTRADIOL <int>,
## #   EXELON <int>, FAMOTIDINE <int>, FELODIPINE.ER <int>,
## #   FENOFIBRATE <int>, FENTANYL <int>, FINASTERIDE <int>,
## #   FLOVENT.HFA <int>, FLUCONAZOLE <int>, FLUOXETINE.HCL <int>,
## #   FLUTICASONE.PROPIONATE <int>, FUROSEMIDE <int>, GABAPENTIN <int>,
## #   GEMFIBROZIL <int>, GLIMEPIRIDE <int>, GLIPIZIDE <int>,
## #   GLIPIZIDE.ER <int>, GLIPIZIDE.XL <int>, GLYBURIDE <int>,
## #   HALOPERIDOL <int>, HUMALOG <int>, HYDRALAZINE.HCL <int>,
## #   HYDROCHLOROTHIAZIDE <int>, HYDROCODONE.ACETAMINOPHEN <int>, ...
```

```
dim(Opioid.PrescribersDF)
```

```
## [1] 14688   256
```

```
VSRR_Provisional_Drug_Overdose_Death_Counts<-read.csv('VSRR_Provisional_Drug_Overdose_Death_Counts.csv')
VSRR_Quarterly_provisional_estimates_for_selected_indicators_of_mortality <- read_csv("~/R_STUDIO_FALL_
```

```
## Parsed with column specification:
## cols(
##   `Year and Quarter` = col_character(),
```

```
##   Indicator = col_character(),
##   `Time Period` = col_character(),
##   `Rate Type` = col_character(),
##   Rate = col_double(),
##   Unit = col_character(),
##   Significant = col_character()
## )
```

```
str(VSRR_Provisional_Drug_Overdose_Death_Counts)
```

```
## 'data.frame':    6214 obs. of  10 variables:
##  $ State                        : Factor w/ 53 levels "AK","AL","AR",..: 22 34 48 1 1 1 1 1 1 1 ...
##  $ State.Name                   : Factor w/ 53 levels "Alabama","Alaska",..: 20 29 48 2 2 2 2 2 2 2
##  $ Year                         : int  2015 2016 2016 2016 2015 2015 2015 2015 2015 2015 ...
##  $ Month                        : Factor w/ 12 levels "April","August",..: 10 6 6 9 5 4 8 1 9 7 ...
##  $ Period                       : Factor w/ 1 level "12 month-ending": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Indicator                    : Factor w/ 9 levels "Cocaine (T40.5)",..: 1 3 5 5 5 5 5 5 5 5 ...
##  $ Data.Value                   : num  33 59 5715 4241 4034 ...
##  $ Percent.Complete             : Factor w/ 5 levels "100","94.2","95",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Percent.Pending.Investigation: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Footnote                     : Factor w/ 2 levels "","*Likely underreported due to incomplete data
```

```
tail(VSRR_Quarterly_provisional_estimates_for_selected_indicators_of_mortality)
```

```
## # A tibble: 6 x 7
##   `Year and Quarter`              Indicator                 `Time Period`
##                <chr>                  <chr>                        <chr>
## 1            2015 Q4 Unintentional injuries 12 months ending with quarter
## 2            2016 Q1 Unintentional injuries 12 months ending with quarter
## 3            2016 Q2 Unintentional injuries 12 months ending with quarter
## 4            2016 Q3 Unintentional injuries 12 months ending with quarter
## 5            2016 Q4 Unintentional injuries 12 months ending with quarter
## 6            2017 Q1 Unintentional injuries 12 months ending with quarter
## # ... with 4 more variables: `Rate Type` <chr>, Rate <dbl>, Unit <chr>,
## #   Significant <chr>
```

```
# head(VSR13.14.drugoverdosedeaths)
```

```
parallel <- select(VSRR_Quarterly_provisional_estimates_for_selected_indicators_of_mortality, c("Year a
```

```
# library(lattice)
# parallelplot(sepopioidRX[c("2015 Q1","2017 Q1"),], horizontal.axis=FALSE)
```

```
Allcauses_Drugoverdose_indicators_estimates <- VSRR_Quarterly_provisional_estimates_for_selected_indica
```

```
Allcauses_Drugoverdose_indicators_estimates<- Allcauses_Drugoverdose_indicators_estimates %>%
  separate(`Year and Quarter`, into=c("Year", "Q"), sep=" ")
Allcauses_Drugoverdose_indicators_estimates$Year <- as.numeric(Allcauses_Drugoverdose_indicators_estima
```

```
Allcauses_Drugoverdose_indicators_estimates <- Allcauses_Drugoverdose_indicators_estimates[Allcauses_Dr
```

```
ODdeaths_13.14 <- OD_Multiple_Cause_of_Death_1999_2014_v1_1[OD_Multiple_Cause_of_Death_1999_2014_v1_1$Y
```

```
head(ODdeaths_13.14)
```

```
## # A tibble: 6 x 8
##     State  Year Deaths Population `Crude Rate`
##     <chr> <int>  <chr>      <int>        <chr>
```

```
## 1 Alabama  2013    175    4833722          3.6
## 2 Alabama  2014    282    4849377          5.8
## 3  Alaska  2013     69     735132          9.4
## 4  Alaska  2014     79     736732         10.7
## 5 Arizona  2013    545    6626624          8.2
## 6 Arizona  2014    616    6731484          9.2
## # ... with 3 more variables: `Crude Rate Lower 95% Confidence
## #   Interval` <chr>, `Crude Rate Upper 95% Confidence Interval` <chr>,
## #   `Prescriptions Dispensed by US Retailers in that year
## #   (millions)` <int>
```

```
# colnames(ACS_13_5YR_B07012_Poverty)
colkeep.pov <- c("Geography","Estimate; Total:","Estimate; Total: - Below 100 percent of the poverty lev
ACS_Poverty_Estimates13 <- select(ACS_13_5YR_B07012_Poverty, colkeep.pov)
```

```
ACS14_Poverty <- select(ACS_14_5YR_B07012_PovertyState, colkeep.pov)
```

```
list1 <- 1:51
list13 <- rep(2013,length(list1))
list14 <- rep(2014,length(list1))
ACS_Poverty_Estimates13$Year <- c(list13)
ACS14_Poverty$Year <- c(list14)
```

```
head(ACS_Poverty_Estimates13, 2)
```

```
## # A tibble: 2 x 6
##   Geography `Estimate; Total:`
##       <chr>              <int>
## 1   Alabama            4628774
## 2    Alaska             693195
## # ... with 4 more variables: `Estimate; Total: - Below 100 percent of the
## #   poverty level` <int>, `Estimate; Total: - 100 to 149 percent of the
## #   poverty level` <int>, `Estimate; Total: - At or above 150 percent of
## #   the poverty level` <int>, Year <dbl>
```

```
head(ACS14_Poverty)
```

```
## # A tibble: 6 x 6
##    Geography `Estimate; Total:`
##        <chr>              <int>
## 1    Alabama            4644377
## 2     Alaska             700504
## 3    Arizona            6331311
## 4   Arkansas            2828552
## 5 California           36872560
## 6   Colorado            5016535
## # ... with 4 more variables: `Estimate; Total: - Below 100 percent of the
## #   poverty level` <int>, `Estimate; Total: - 100 to 149 percent of the
## #   poverty level` <int>, `Estimate; Total: - At or above 150 percent of
## #   the poverty level` <int>, Year <dbl>
```

```
total13.14.Poverty <- rbind(ACS_Poverty_Estimates13, ACS14_Poverty)
```

```
colnames(total13.14.Poverty) <- c("State", "Total Poverty", "Below 100 percent of the poverty level", "
total13.14.Poverty <- total13.14.Poverty[order(total13.14.Poverty$State),]
tail(total13.14.Poverty)
```

```
## # A tibble: 6 x 6
##           State `Total Poverty` `Below 100 percent of the poverty level`
##           <chr>           <int>                                     <int>
## 1 West Virginia         1781742                                    315861
## 2 West Virginia         1781386                                    321003
## 3     Wisconsin         5489893                                    710014
## 4     Wisconsin         5506923                                    724480
## 5       Wyoming          549007                                     62147
## 6       Wyoming          554316                                     63826
## # ... with 3 more variables: `100 to 149 percent of the poverty
## #   level` <int>, `At or above 150 percent of the poverty level` <int>,
## #   Year <dbl>
```

```
dim(total13.14.Poverty)
```

```
## [1] 102   6
```

```
Poverty.ODeaths.13.15 <- merge(total13.14.Poverty, ODdeaths_13.14, by=c("State","Year"))
```

```
head(Poverty.ODeaths.13.15, 2)
```

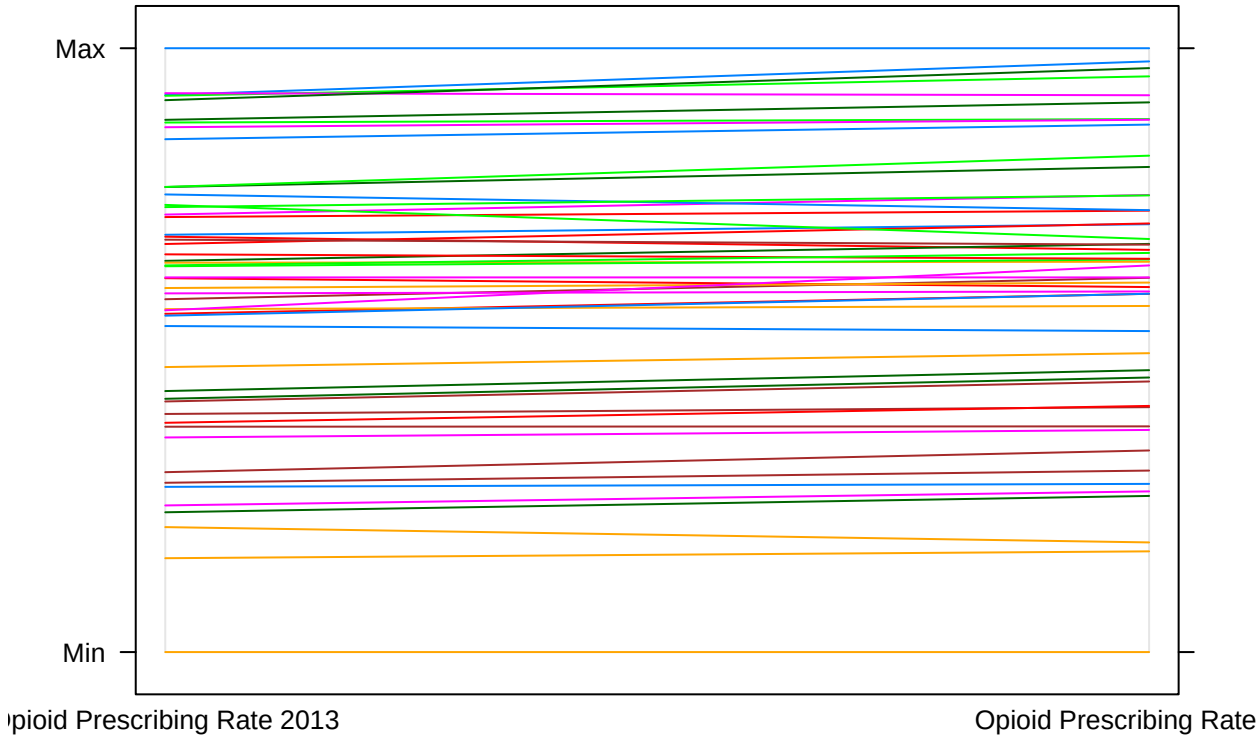
```
##     State Year Total Poverty Below 100 percent of the poverty level
## 1 Alabama 2013       4628774                                853751
## 2 Alabama 2014       4644377                                871902
##   100 to 149 percent of the poverty level
## 1                                  516964
## 2                                  519257
##   At or above 150 percent of the poverty level Deaths Population
## 1                                      3258059    175    4833722
## 2                                      3253218    282    4849377
##   Crude Rate Crude Rate Lower 95% Confidence Interval
## 1        3.6                                     3.1
## 2        5.8                                     5.1
##   Crude Rate Upper 95% Confidence Interval
## 1                                      4.2
## 2                                      6.5
##   Prescriptions Dispensed by US Retailers in that year (millions)
## 1                                                             207
## 2                                                             196
```

```
setnames(Poverty.ODeaths.13.15, "Crude Rate", "Crude Rate Deaths")
```

```
colnames(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014)
```

```
##  [1] "State"
##  [2] "State Abbreviation"
##  [3] "State FIPS"
##  [4] "Part D Prescribers 2013"
##  [5] "Opioid Claims 2013"
##  [6] "Overall Claims 2013"
##  [7] "Opioid Prescribing Rate 2013"
##  [8] "Part D Prescribers 2014"
##  [9] "Opioid Claims 2014"
## [10] "Overall Claims 2014"
## [11] "Opioid Prescribing Rate 2014"
## [12] "Percentage Point Difference in Opioid Prescribing Rate"
```

```
library(lattice)
parallelplot(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014[, c("Opioid Prescribing Rate 2013","
```




```
PartD.Opioid.13 <- select(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014, c("State", "State Abbr
# list1 <- 1:51
# list13 <- rep(2013,length(list1))
# list14 <- rep(2014,length(list1))
PartD.Opioid.13$Year <- c(list13)
PartD.Opioid.14 <- select(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014, c("State", "State Abbr
```

```
PartD.Opioid.14$Year <- c(list14)
new.names.partD <- c("State", "Abbrev", "Part D Prescribers", "Opioid Claims", "Overall Claims", "Opioi
colnames(PartD.Opioid.13) <- new.names.partD
colnames(PartD.Opioid.14) <- new.names.partD
```

```
total13.14.Prescribers <- rbind(PartD.Opioid.13, PartD.Opioid.14)
```

```
Poverty.ODeaths.Prescribers.13.14 <- merge(total13.14.Prescribers, Poverty.ODeaths.13.15, by=c("State",
```

```
head(Poverty.ODeaths.Prescribers.13.14)
```

```
##     State Year Abbrev Part D Prescribers Opioid Claims Overall Claims
## 1 Alabama 2013     AL              12820       2260284       29160952
## 2 Alabama 2014     AL              13100       2267090       28852731
## 3  Alaska 2013     AK               2275         86517        1281057
## 4  Alaska 2014     AK               2293         93606        1384451
## 5 Arizona 2013     AZ              20542       1545138       22126421
## 6 Arizona 2014     AZ              21253       1639782       23454968
##   Opioid Prescribing Rate Total Poverty
## 1                7.751064       4628774
## 2                7.857454       4644377
```

```
## 3                6.753564        693195
## 4                6.761236        700504
## 5                6.983226       6250872
## 6                6.991193       6331311
##   Below 100 percent of the poverty level
## 1                                853751
## 2                                871902
## 3                                 67553
## 4                                 69873
## 5                               1107636
## 6                               1145730
##   100 to 149 percent of the poverty level
## 1                                 516964
## 2                                 519257
## 3                                  55420
## 4                                  55966
## 5                                 674872
## 6                                 690020
##   At or above 150 percent of the poverty level Deaths Population
## 1                                      3258059    175    4833722
## 2                                      3253218    282    4849377
## 3                                       570222     69     735132
## 4                                       574665     79     736732
## 5                                      4468364    545    6626624
## 6                                      4495561    616    6731484
##   Crude Rate Deaths Crude Rate Lower 95% Confidence Interval
## 1               3.6                                     3.1
## 2               5.8                                     5.1
## 3               9.4                                     7.3
## 4              10.7                                     8.5
## 5               8.2                                     7.5
## 6               9.2                                     8.4
##   Crude Rate Upper 95% Confidence Interval
## 1                                      4.2
## 2                                      6.5
## 3                                     11.9
## 4                                     13.4
## 5                                      8.9
## 6                                      9.9
##   Prescriptions Dispensed by US Retailers in that year (millions)
## 1                                                             207
## 2                                                             196
## 3                                                             207
## 4                                                             196
## 5                                                             207
## 6                                                             196
```

```
colnames(Poverty.ODeaths.Prescribers.13.14)
```

```
##  [1] "State"
##  [2] "Year"
##  [3] "Abbrev"
##  [4] "Part D Prescribers"
##  [5] "Opioid Claims"
##  [6] "Overall Claims"
```

```
##  [7] "Opioid Prescribing Rate"
##  [8] "Total Poverty"
##  [9] "Below 100 percent of the poverty level"
## [10] "100 to 149 percent of the poverty level"
## [11] "At or above 150 percent of the poverty level"
## [12] "Deaths"
## [13] "Population"
## [14] "Crude Rate Deaths"
## [15] "Crude Rate Lower 95% Confidence Interval"
## [16] "Crude Rate Upper 95% Confidence Interval"
## [17] "Prescriptions Dispensed by US Retailers in that year (millions)"
```

```
Poverty.ODeaths.Prescribers.13.14$Opioid.Overall <- (Poverty.ODeaths.Prescribers.13.14$`Opioid Claims`/
```

```
Poverty.ODeaths.Prescribers.13.14$Ratio.Prescribers.Popln <- round(100000*(Poverty.ODeaths.Prescribers.
```

```
head(Poverty.ODeaths.Prescribers.13.14)
```

```
##     State Year Abbrev Part D Prescribers Opioid Claims Overall Claims
## 1 Alabama 2013     AL              12820       2260284       29160952
## 2 Alabama 2014     AL              13100       2267090       28852731
## 3  Alaska 2013     AK               2275         86517        1281057
## 4  Alaska 2014     AK               2293         93606        1384451
## 5 Arizona 2013     AZ              20542       1545138       22126421
## 6 Arizona 2014     AZ              21253       1639782       23454968
##   Opioid Prescribing Rate Total Poverty
## 1                7.751064       4628774
## 2                7.857454       4644377
## 3                6.753564        693195
## 4                6.761236        700504
## 5                6.983226       6250872
## 6                6.991193       6331311
##   Below 100 percent of the poverty level
## 1                                 853751
## 2                                 871902
## 3                                  67553
## 4                                  69873
## 5                                1107636
## 6                                1145730
##   100 to 149 percent of the poverty level
## 1                                  516964
## 2                                  519257
## 3                                   55420
## 4                                   55966
## 5                                  674872
## 6                                  690020
##   At or above 150 percent of the poverty level Deaths Population
## 1                                      3258059    175    4833722
## 2                                      3253218    282    4849377
## 3                                       570222     69     735132
## 4                                       574665     79     736732
## 5                                      4468364    545    6626624
## 6                                      4495561    616    6731484
##   Crude Rate Deaths Crude Rate Lower 95% Confidence Interval
## 1               3.6                                     3.1
```

```
## 2              5.8                                        5.1
## 3              9.4                                        7.3
## 4             10.7                                        8.5
## 5              8.2                                        7.5
## 6              9.2                                        8.4
##   Crude Rate Upper 95% Confidence Interval
## 1                                      4.2
## 2                                      6.5
## 3                                     11.9
## 4                                     13.4
## 5                                      8.9
## 6                                      9.9
##   Prescriptions Dispensed by US Retailers in that year (millions)
## 1                                                             207
## 2                                                             196
## 3                                                             207
## 4                                                             196
## 5                                                             207
## 6                                                             196
##   Opioid.Overall Ratio.Prescribers.Popln
## 1     0.07751064                    265
## 2     0.07857454                    270
## 3     0.06753564                    309
## 4     0.06761236                    311
## 5     0.06983226                    310
## 6     0.06991193                    316
```

```
opioidanalgesicRXtibble <- as_tibble(Opioid_analgesic_prescriptions_dispensed_from_US_retail_pharmacies_
```

```
head(opioidanalgesicRXtibble, 4)
```

```
## # A tibble: 4 x 13
##   Quarter Hydrocodone Oxycodone Tramadol Morphine Fentanyl `H+O+T+M+F`
##     <chr>       <int>     <int>    <int>    <int>    <int>       <int>
## 1 Q4 2009    30831801  12598334  6475986  1731324  1266443    52903888
## 2 Q1 2010    30692922  12676513  6429833  1727199  1240702    52767169
## 3 Q2 2010    31554204  13282829  6694127  1792189  1274043    54597392
## 4 Q3 2010    32000678  13748219  6811394  1833943  1283203    55677437
## # ... with 6 more variables: `All Opioid Analgesics` <int>, `ER/LA Opioid
## #   Analgesics` <int>, `Yearly totals (All Opioid Analgesics)` <chr>,
## #   `Yearly totals (ER/LA Opioid Analgesics)` <chr>, `Yearly totals
## #   (H+O+T+M+F)` <chr>, `Yearly totals (H+O)` <chr>
```

```
sepopioidRX <- opioidanalgesicRXtibble %>%
  separate(Quarter, into=c("Q", "Year"), sep=" ")
```

```
sepopioidRX$Year <- as.numeric(sepopioidRX$Year)
```

```
head(sepopioidRX, 2)
```

```
## # A tibble: 2 x 14
##       Q  Year Hydrocodone Oxycodone Tramadol Morphine Fentanyl `H+O+T+M+F`
##   <chr> <dbl>       <int>     <int>    <int>    <int>    <int>       <int>
## 1    Q4  2009    30831801  12598334  6475986  1731324  1266443    52903888
## 2    Q1  2010    30692922  12676513  6429833  1727199  1240702    52767169
## # ... with 6 more variables: `All Opioid Analgesics` <int>, `ER/LA Opioid
```

```
## #   Analgesics` <int>, `Yearly totals (All Opioid Analgesics)` <chr>,
## #   `Yearly totals (ER/LA Opioid Analgesics)` <chr>, `Yearly totals
## #   (H+O+T+M+F)` <chr>, `Yearly totals (H+O)` <chr>
```

```
sepopioidRX$Num <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23)
```

```
sepopioidRX <-as.data.frame(sepopioidRX)
```

```
sepopioidRX
```

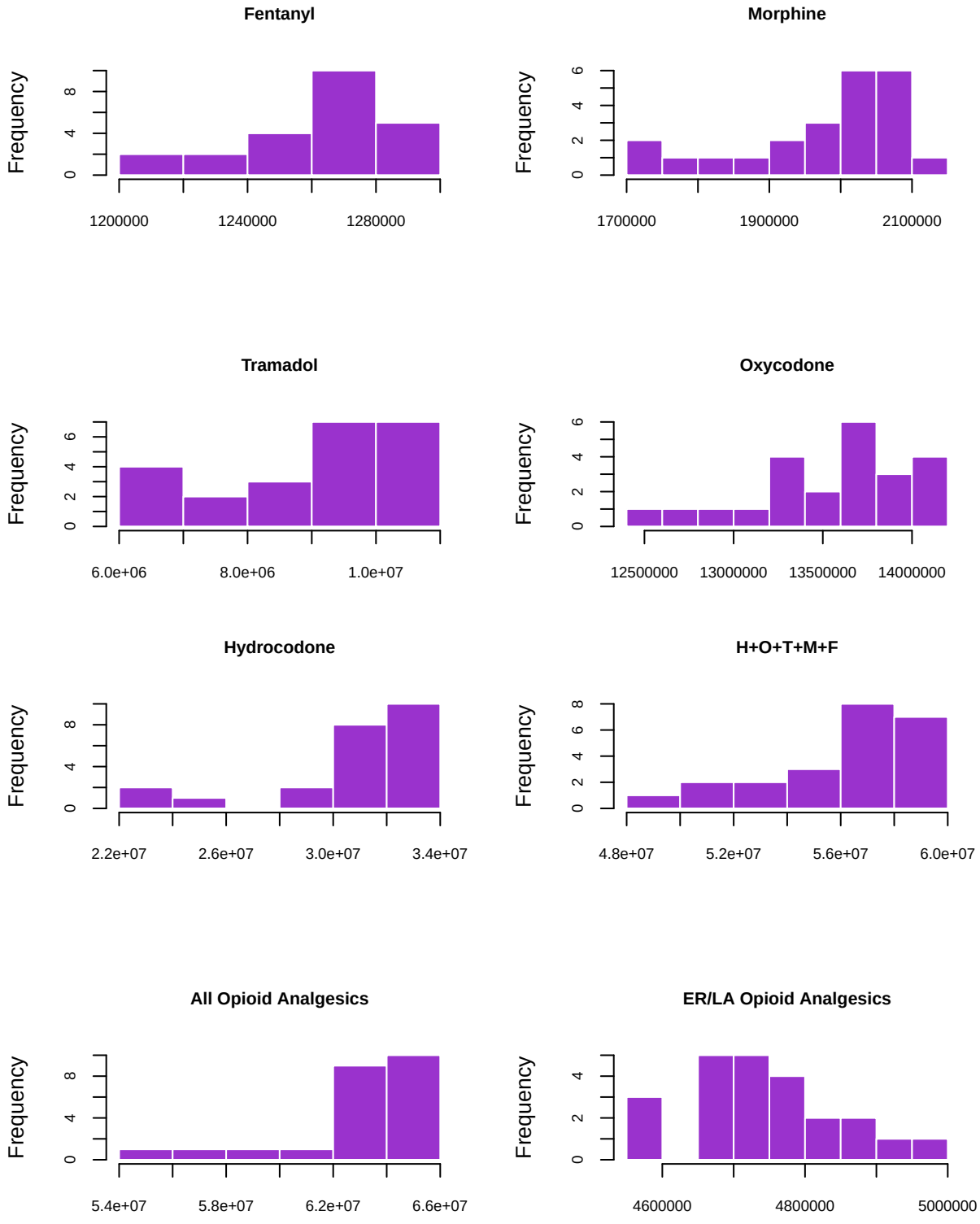
```
##     Q Year Hydrocodone Oxycodone Tramadol Morphine Fentanyl H+O+T+M+F
## 1  Q4 2009    30831801  12598334  6475986  1731324  1266443  52903888
## 2  Q1 2010    30692922  12676513  6429833  1727199  1240702  52767169
## 3  Q2 2010    31554204  13282829  6694127  1792189  1274043  54597392
## 4  Q3 2010    32000678  13748219  6811394  1833943  1283203  55677437
## 5  Q4 2010    32239057  13825106  7453897  1889202  1287753  56695015
## 6  Q1 2011    32497676  13697064  7934498  1904519  1271533  57305290
## 7  Q2 2011    32906191  13972641  8194032  1932387  1290096  58295347
## 8  Q3 2011    32970249  14129677  8461645  1959390  1296659  58817620
## 9  Q4 2011    32557588  14015327  8426827  1985884  1292906  58278532
## 10 Q1 2012    32821380  13676982  9085356  1967390  1256587  58807695
## 11 Q2 2012    32883231  13799118  9373129  2014571  1270489  59340538
## 12 Q3 2012    32621466  13714150  9556588  2021396  1267369  59180969
## 13 Q4 2012    32429704  13810577  9753897  2058168  1273738  59326084
## 14 Q1 2013    30848355  13134306  9488158  2005937  1233775  56710531
## 15 Q2 2013    31097912  13251033  9913793  2032962  1260428  57556128
## 16 Q3 2013    31081430  13310718 10096891  2060817  1267488  57817344
## 17 Q4 2013    30629191  13405070 10327890  2073818  1261975  57697944
## 18 Q1 2014    29299051  12965080 10156105  2010032  1215947  55646215
## 19 Q2 2014    29932633  13362307 10725262  2069597  1249482  57339281
## 20 Q3 2014    30088966  13641359 10396635  2089791  1256113  57472864
## 21 Q4 2014    24111897  14074835 10435370  2115022  1261672  51998796
## 22 Q1 2015    22881381  13595329  9954737  2041447  1202643  49675537
## 23 Q2 2015    23384486  14066532 10182119  2091995  1232571  50957703
##    All Opioid Analgesics ER/LA Opioid Analgesics
## 1                62804834                 4808919
## 2                62407792                 4727239
## 3                64359386                 4886616
## 4                65385955                 4960528
## 5                64915764                 4900668
## 6                63667268                 4750490
## 7                64655990                 4778769
## 8                65093125                 4804171
## 9                64535529                 4852112
## 10               64992656                 4674487
## 11               65346051                 4702975
## 12               65030875                 4689927
## 13               65163879                 4763795
## 14               62344354                 4593200
## 15               63108593                 4660126
## 16               63279780                 4708521
## 17               63101698                 4732618
## 18               60833351                 4559309
## 19               62555737                 4689143
## 20               62699948                 4732605
```

```
## 21              58454163                 4791578
## 22              55852248                 4568297
## 23              57274927                 4660385
##    Yearly totals (All Opioid Analgesics)
## 1                              No data
## 2                            257068897
## 3                            257068897
## 4                            257068897
## 5                            257068897
## 6                            257951912
## 7                            257951912
## 8                            257951912
## 9                            257951912
## 10                           260533461
## 11                           260533461
## 12                           260533461
## 13                           260533461
## 14                           251834425
## 15                           251834425
## 16                           251834425
## 17                           251834425
## 18                           244543199
## 19                           244543199
## 20                           244543199
## 21                           244543199
## 22                             No data
## 23                             No data
##    Yearly totals (ER/LA Opioid Analgesics) Yearly totals (H+O+T+M+F)
## 1                                 No data                   No data
## 2                                19475051                 219737013
## 3                                19475051                 219737013
## 4                                19475051                 219737013
## 5                                19475051                 219737013
## 6                                19185542                 232696789
## 7                                19185542                 232696789
## 8                                19185542                 232696789
## 9                                19185542                 232696789
## 10                               18831184                 229781947
## 11                               18831184                 229781947
## 12                               18831184                 229781947
## 13                               18831184                 229781947
## 14                               18694465                 229781947
## 15                               18694465                 229781947
## 16                               18694465                 229781947
## 17                               18694465                 229781947
## 18                               18772635                 222457156
## 19                               18772635                 222457156
## 20                               18772635                 222457156
## 21                               18772635                 222457156
## 22                                No data                   No data
## 23                                No data                   No data
##    Yearly totals (H+O) Num
## 1             No data   1
## 2           180019528   2
```

```
## 3            180019528   3
## 4            180019528   4
## 5            180019528   5
## 6            186746413   6
## 7            186746413   7
## 8            186746413   8
## 9            186746413   9
## 10           176758015  10
## 11           176758015  11
## 12           176758015  12
## 13           176758015  13
## 14           176758015  14
## 15           176758015  15
## 16           176758015  16
## 17           176758015  17
## 18           167476128  18
## 19           167476128  19
## 20           167476128  20
## 21           167476128  21
## 22             No data  22
## 23             No data  23
```

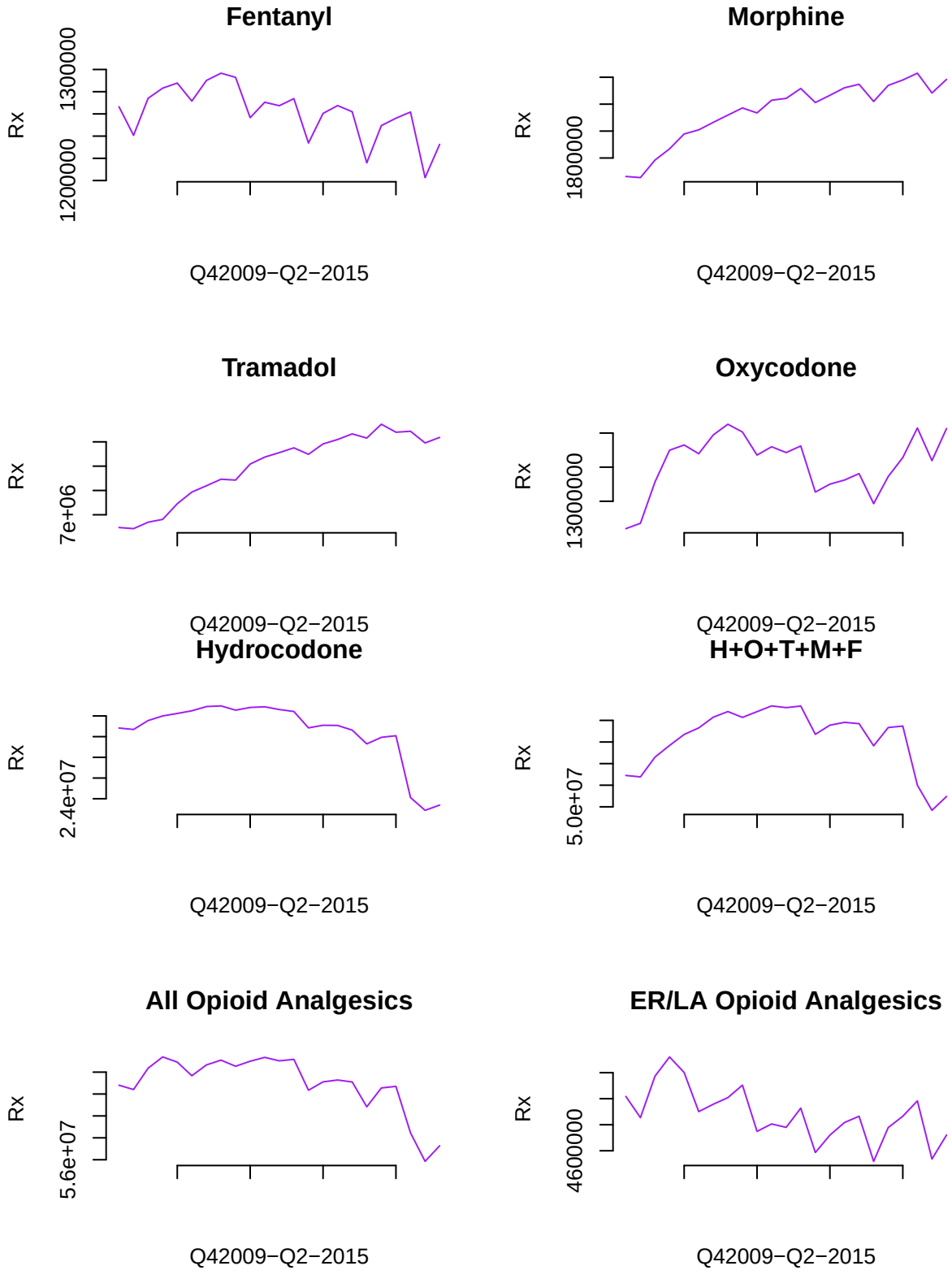
```
library(ggthemes)
library(ggplot2)
```

```
drug_columns <- c("Fentanyl", "Morphine", "Tramadol", "Oxycodone", "Hydrocodone", "H+O+T+M+F", "All Opi
cl <- rainbow(8)
# max_deaths <-max(sepopioidRX[,drug_columns])
par(mfrow=c(2,2))
for (i in 1:length(drug_columns)) {
  hist(as.numeric(sepopioidRX[,drug_columns[i]]), col="darkorchid3", main=drug_columns[i],cex.main=0.8,
}
```




```
par(mfrow=c(2,2))
# Small multiples, lines
for (i in 1:length(drug_columns)) {
  plot(sepopioidRX$Num, sepopioidRX[,drug_columns[i]], type="l", main=drug_columns[i], xlab="Q42009-Q2-
```

```
 axis(1, labels = FALSE)
 axis(side=2, labels=TRUE)
}
```

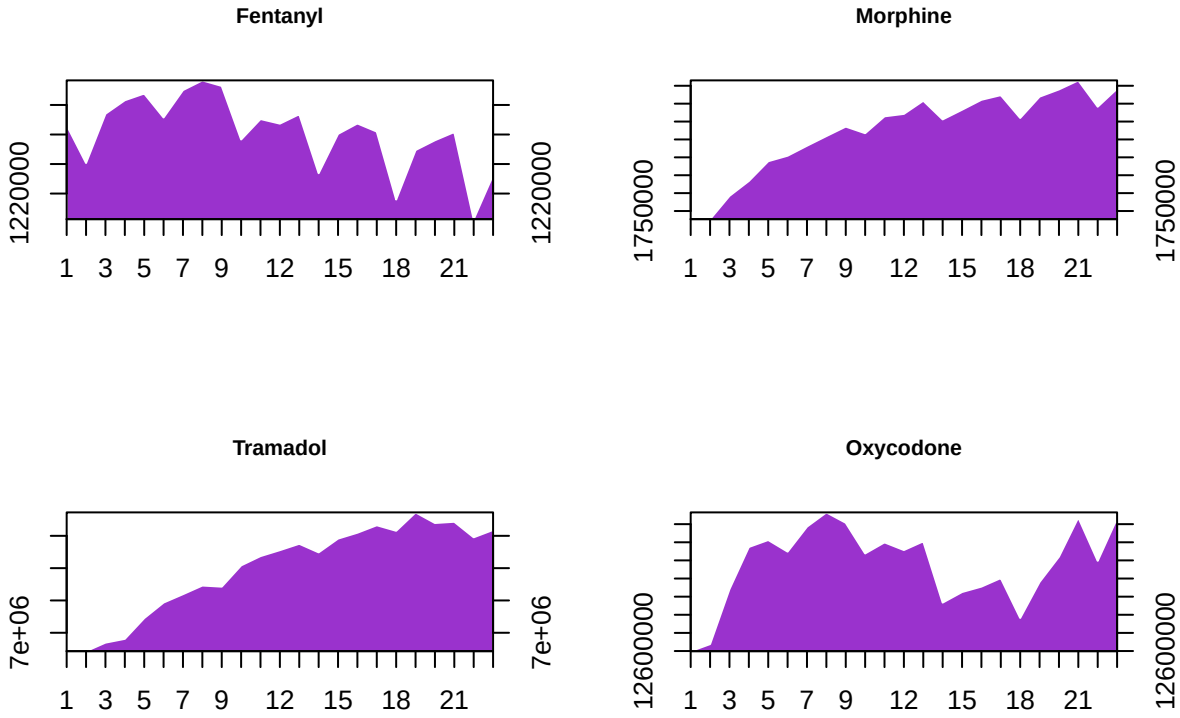


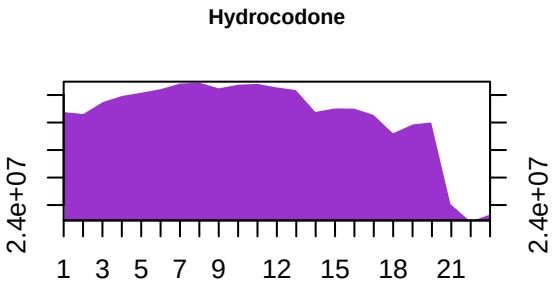
```
library(nlme)
```

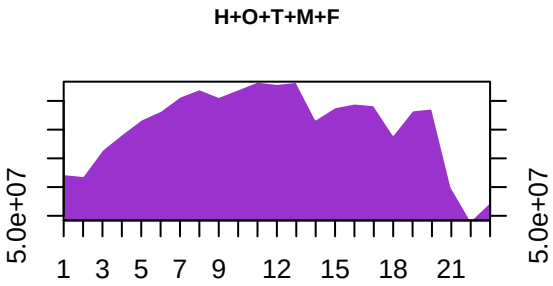
```
##
## Attaching package: 'nlme'
```

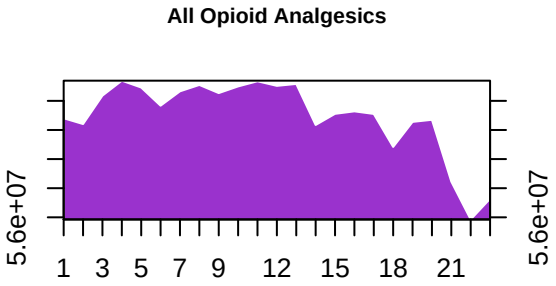
```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

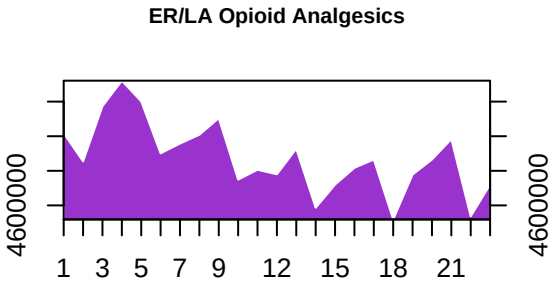
```
library(plotrix)
par(mfrow=c(2,2))
for (i in 1:length(drug_columns)) {
  stackpoly(as.numeric(sepopioidRX[,drug_columns[i]], col=cl[i]), col="darkorchid3", main=drug_columns[
}
```



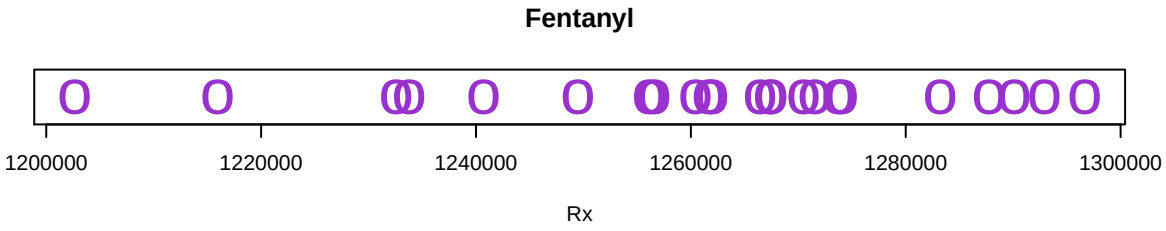


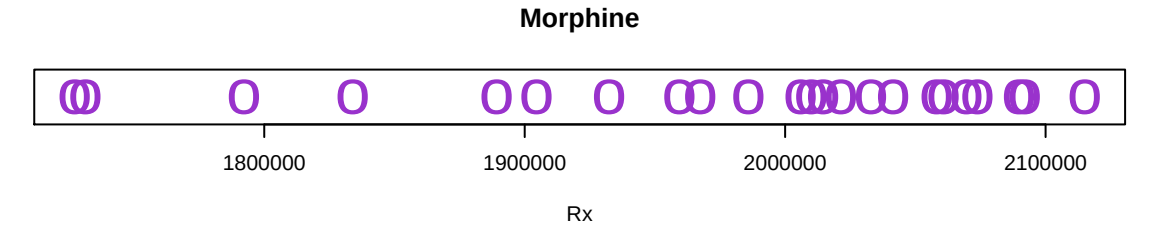


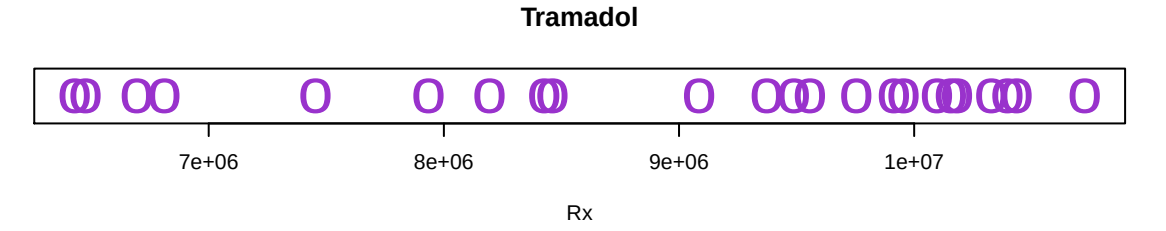


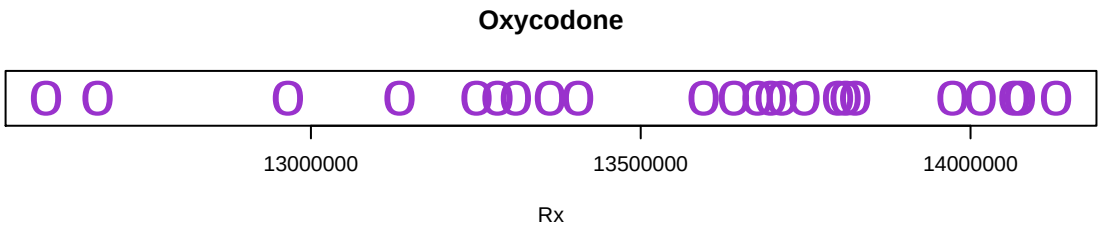


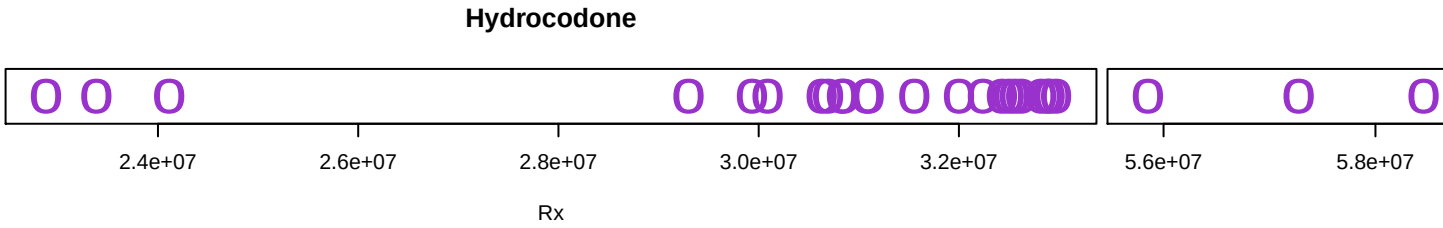

```
library(nlme)
par(mfrow=c(3,1))
for (i in 1:length(drug_columns)) {
 stripchart(sepopioidRX[,drug_columns[i]], main=drug_columns[i], pch='o', col="darkorchid", cex=3, xlab
}
```

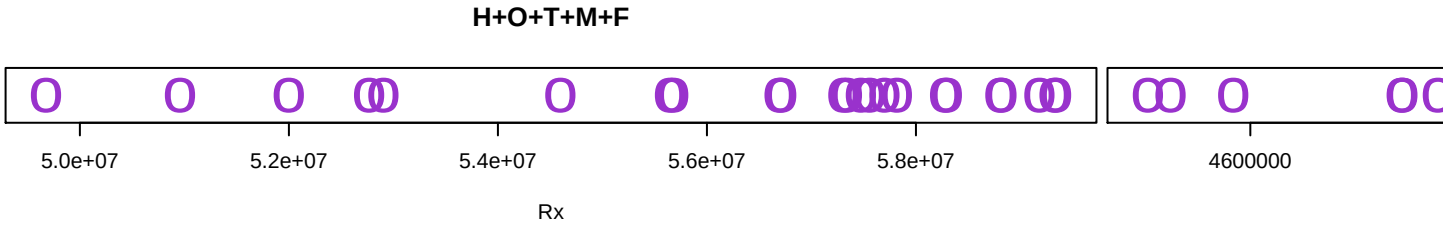




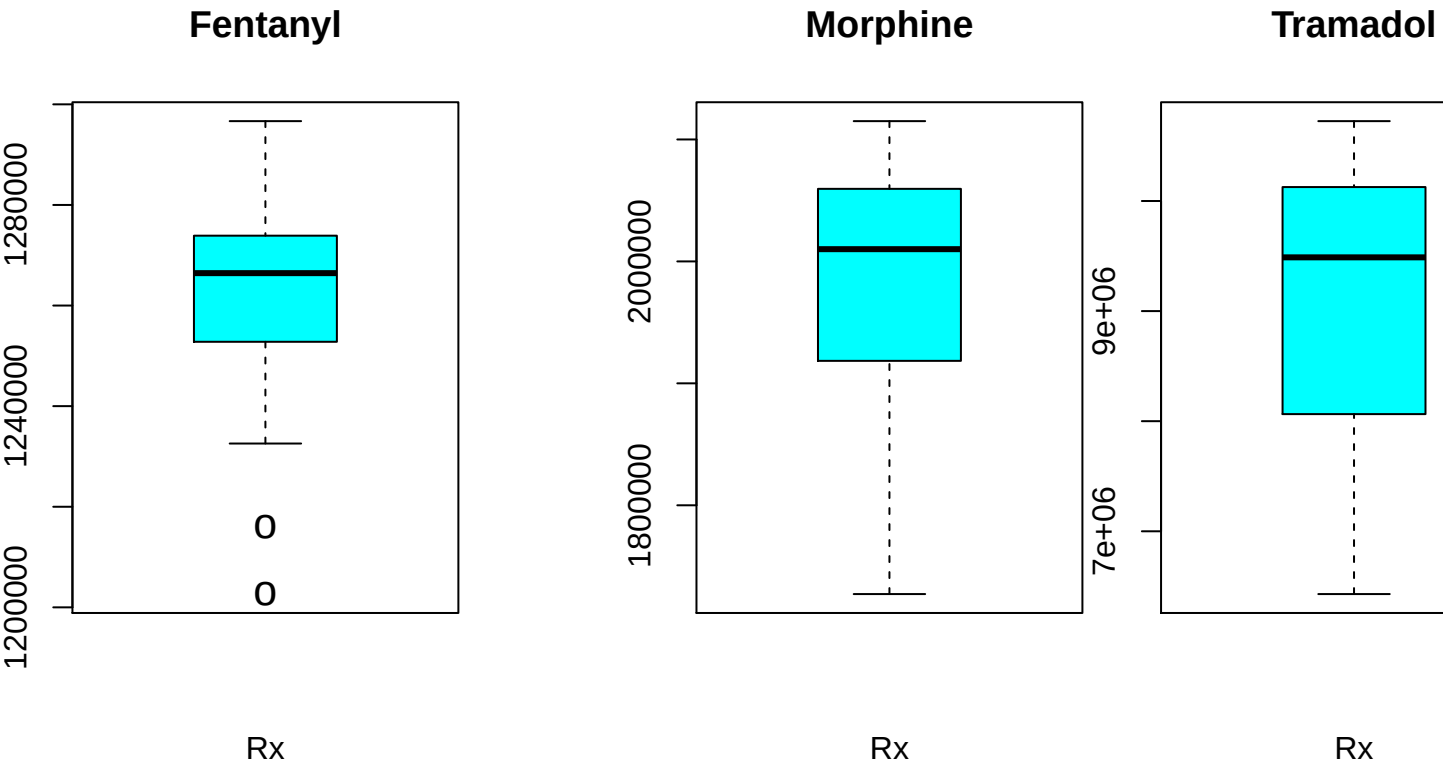


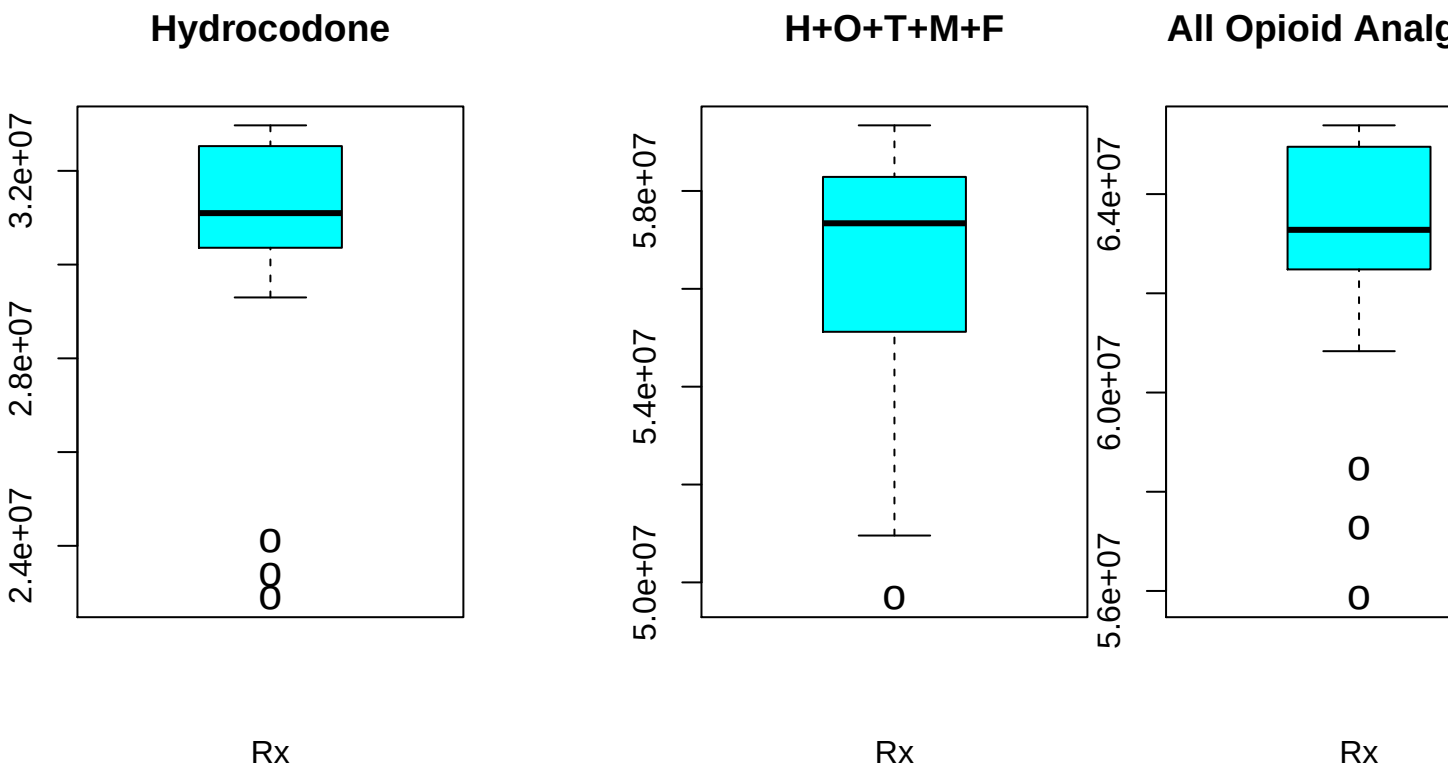




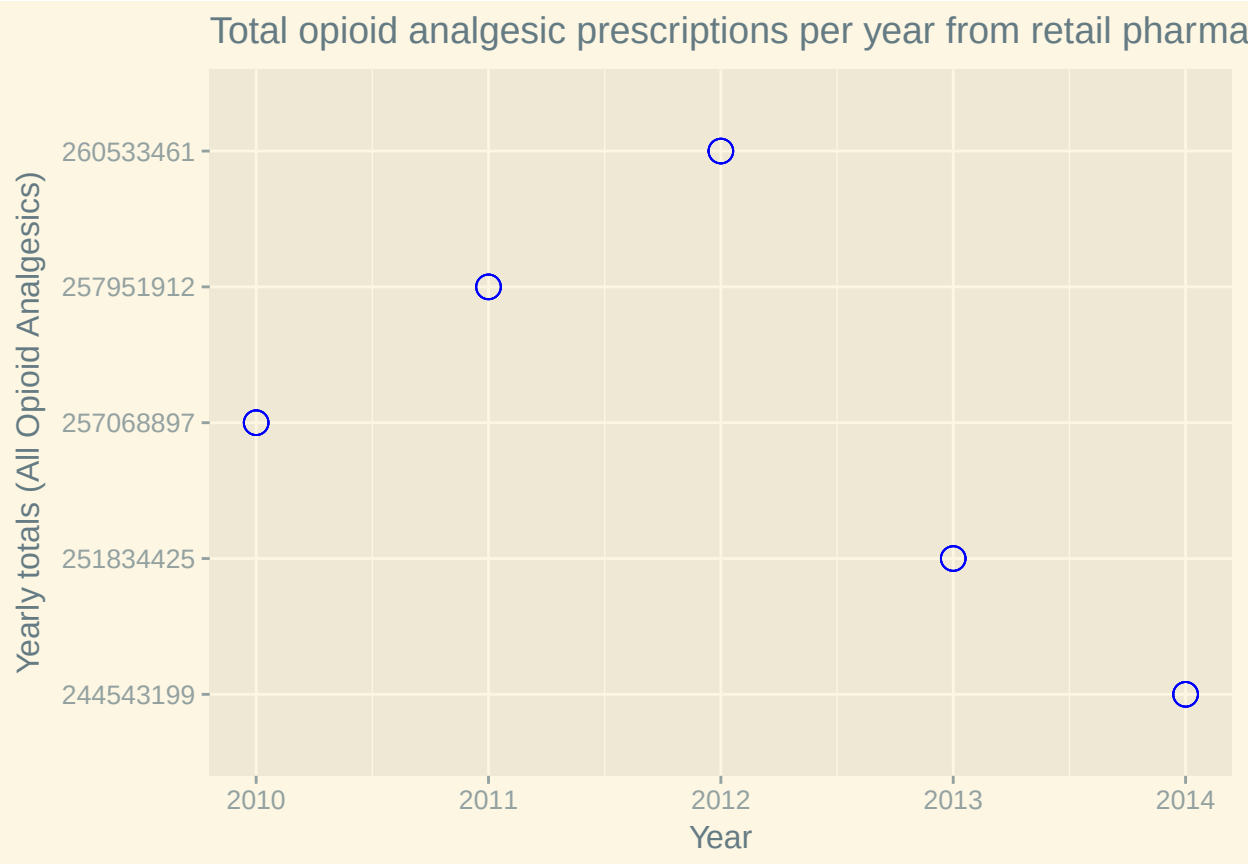



```
par(mfrow=c(1,2))
for (i in 1:length(drug_columns)) {
 boxplot(sepopioidRX[,drug_columns[i]], main=drug_columns[i], pch='o', col="cyan", cex=1.3, xlab="Rx")
}
```

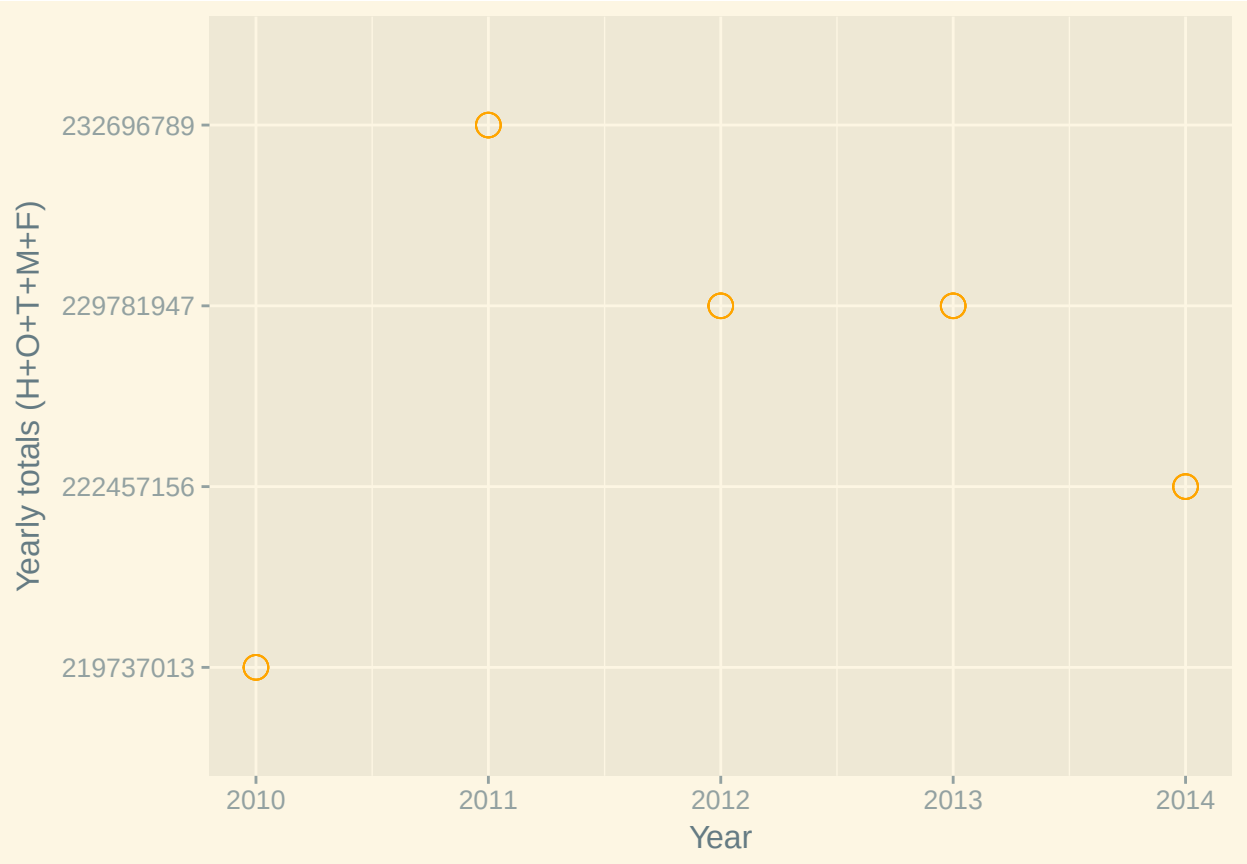





```
sepopioidRXn <- sepopioidRX[sepopioidRX$Year!=2009,]
sepopioidRXn <- sepopioidRXn[sepopioidRXn$Year!=2015,]
p <- ggplot(sepopioidRXn, aes(x=Year)) + theme_solarized_2()
p + labs(title="Total opioid analgesic prescriptions per year from retail pharmacies") + geom_point(aes
```




```
p + geom_point(aes(y=`Yearly totals (H+O+T+M+F)`),size=4,pch=21,col="orange")
```



Large decrease in opioid analgesic Part D claims from 2012 to 2013.

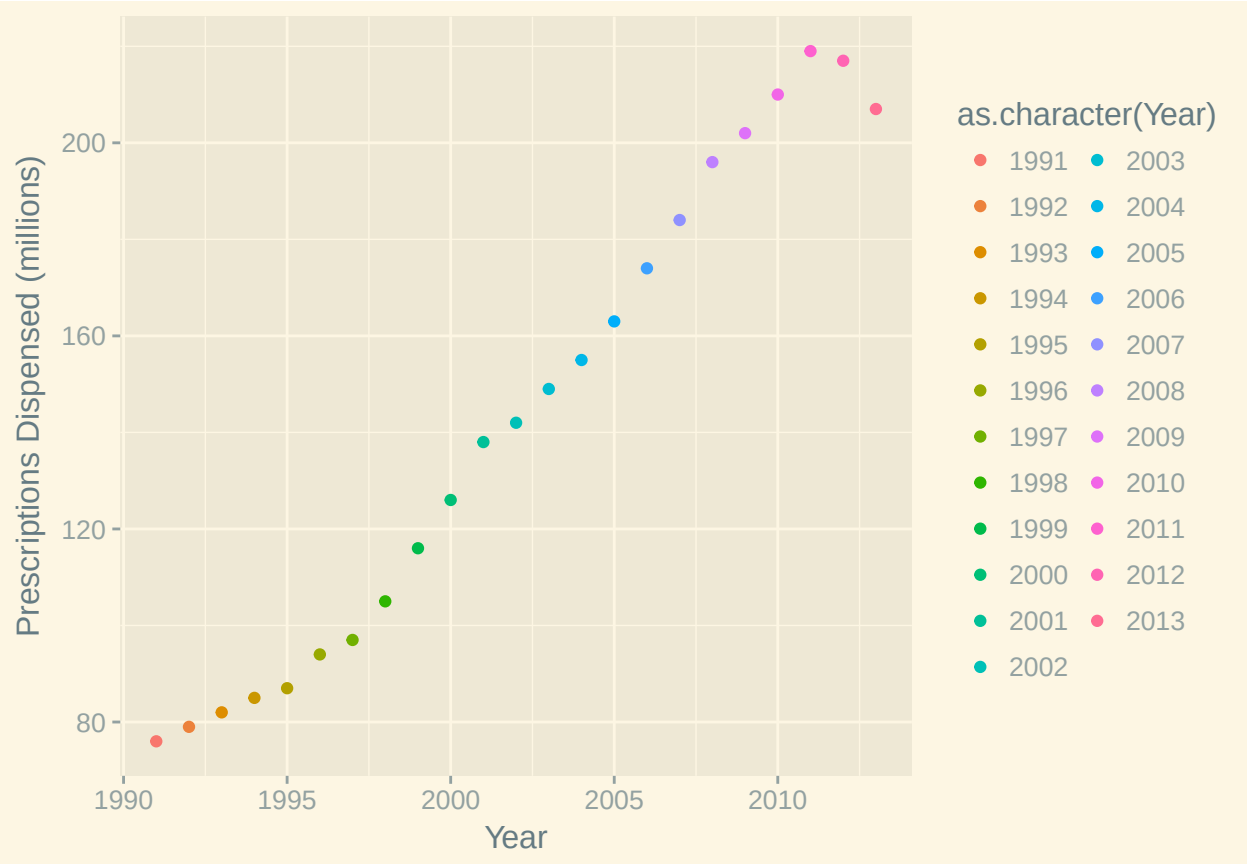
```
# p <- ggplot(sepopioidRX, aes(x=Year, y=`All Opioid Analgesics`, group=Year)) + geom_point(size=3, col
#              ymin = -Inf,ymax = Inf), alpha = 0.25) + labs(title="All Opioid Analgesics per Year (Part
#
# p + facet_grid(. ~Year, scales="fixed") + theme_dark() + scale_color_wsj("colors6", "") + theme(legen
```

What about opoid dispensing from retail pharmacies?

```
opioid_prescriptions_dispensed_us_1991_2013 <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_proj
```

```
## Parsed with column specification:
## cols(
##   Year = col_integer(),
##   `Prescriptions Dispensed (millions)` = col_integer()
## )
```

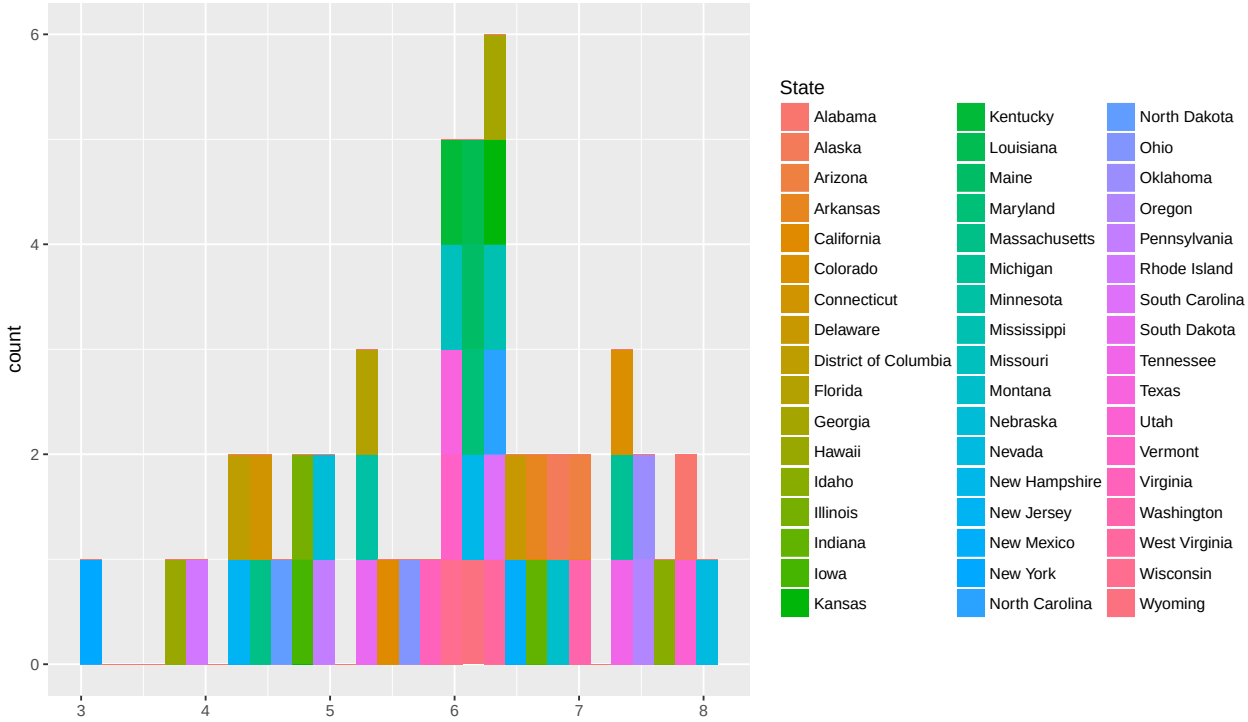
```
ggplot(opioid_prescriptions_dispensed_us_1991_2013, aes(Year, `Prescriptions Dispensed (millions)`)) + g
```



Part D only contains information for those age 65+. This is why I want to look directly at individual prescriber patterns.

```
ggplot(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014) + geom_histogram(aes(Part_D_Opioid_Prescri
```

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



Part_D_Opioid_Prescribing_Change_Geographic_2013_2014$`Opioid Prescribing Rate 2014`

```
Pharmacies <- read_csv("~/R_STUDIO_FALL_2017_PDA/hw2-izzykayu/PDA_project_tweets_machinelearning_IM_MG/
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   FID = col_integer(),
##   ID = col_integer(),
##   ZIP = col_integer(),
##   CONTDATE = col_datetime(format = ""),
##   GEODATE = col_datetime(format = ""),
##   NAICSCODE = col_integer(),
##   X = col_double(),
##   Y = col_double(),
##   NPI = col_integer(),
##   ENT_TYPE = col_integer(),
##   PROVID_12 = col_integer(),
##   PROVID_26 = col_double(),
##   PROVID_27 = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
rx_columns <- c("Fentanyl", "Tramadol", "Oxycodone", "Morphine", "H+O+T+M+F", "All Opioid Analgesics",
# max_deaths <-max(DrugAppearances_Yr_DF[,drug_columns])
# plot(DrugAppearances_Yr_DF$Year, DrugAppearances_Yr_DF$Fentanyl, type="n", ylim=c(0, max_deaths))
# for (i in 1:length(drug_columns)) {
#   lines(DrugAppearances_Yr_DF$Year, DrugAppearances_Yr_DF[,drug_columns[i]])
# }
PharmacyStateCounts <- as.data.frame(multi.fun(Pharmacies$STATE))
```

```
PharmacyStateCounts <- setDT(PharmacyStateCounts, keep.rownames = TRUE)
```

```
# colSums(Filter(is.numeric, people)) makes df of all sums
head(PharmacyStateCounts)
```

```
##    rn freq  percentage
## 1: AK  126 0.200085751
## 2: AL 1256 1.994505582
## 3: AR  680 1.079827863
## 4: AS    1 0.001587982
## 5: AZ 1132 1.797595795
## 6: CA 5802 9.213472441
```

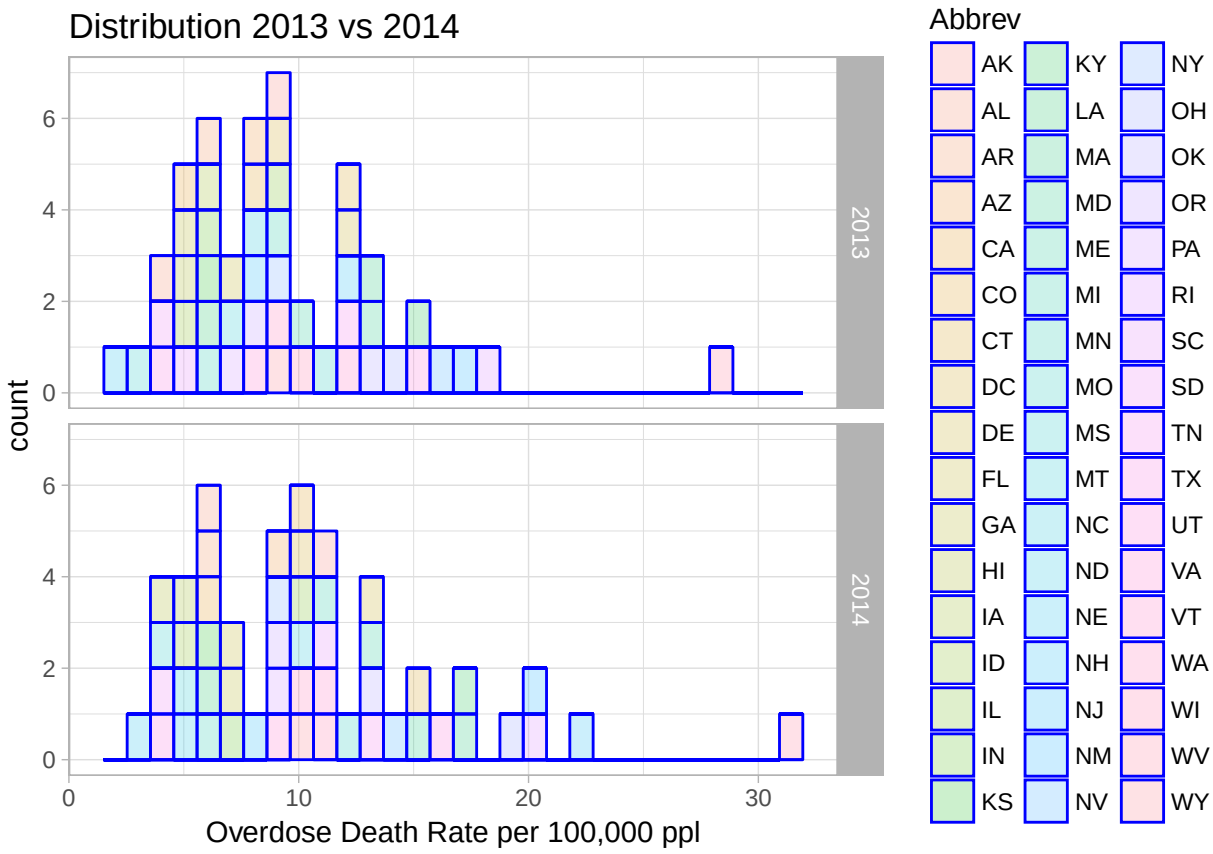
```
library(ggthemes)
# A histogram of bill sizes
columns=c("Crude Rate Deaths", "Opioid Prescribing Rate")

for (i in 1:length(columns)) {
ggplot(data=Poverty.ODeaths.Prescribers.13.14) + geom_histogram(aes(as.numeric(Poverty.ODeaths.Prescrib
}
# Histogram of crude rate deaths, divided by year, colored by state

hp <- ggplot(Poverty.ODeaths.Prescribers.13.14, aes(x=as.numeric(`Crude Rate Deaths`))) + geom_histogra

hp
```

```
## Warning in fun(x, ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

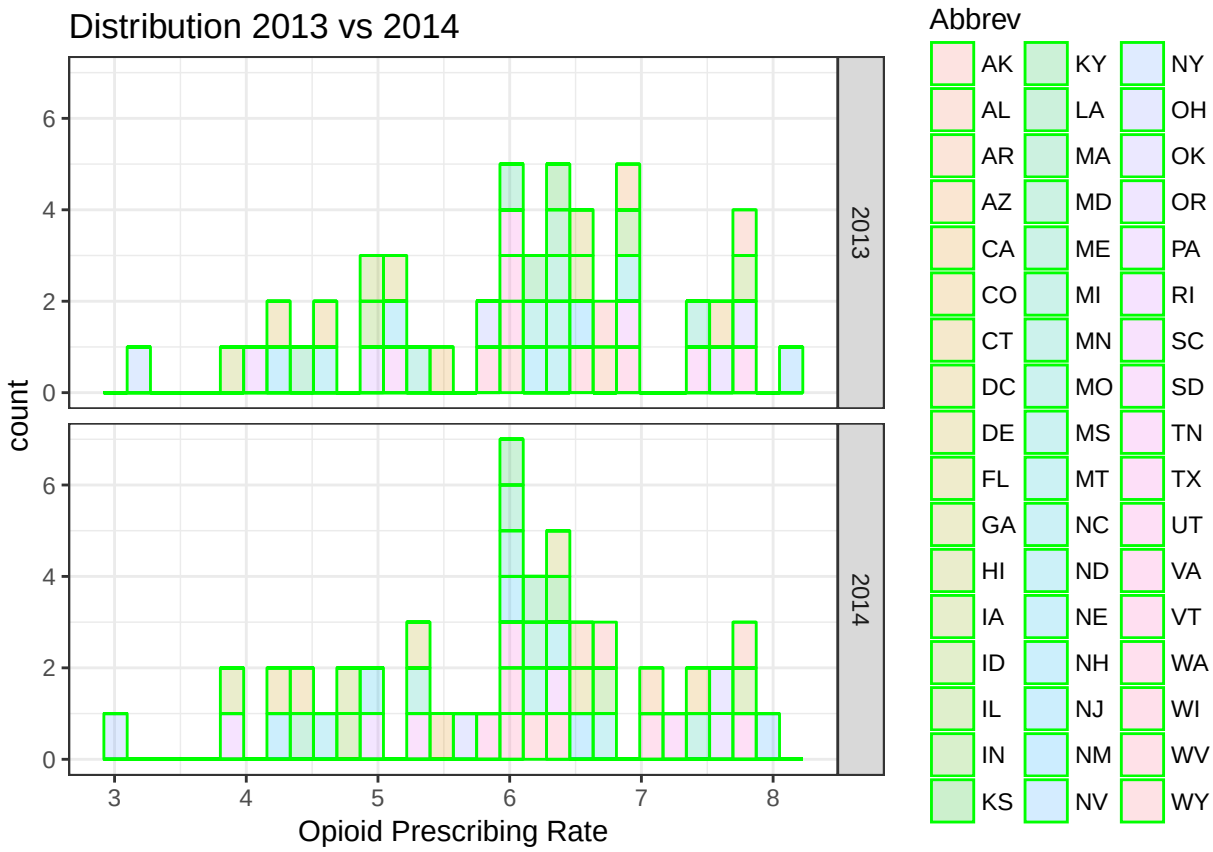
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



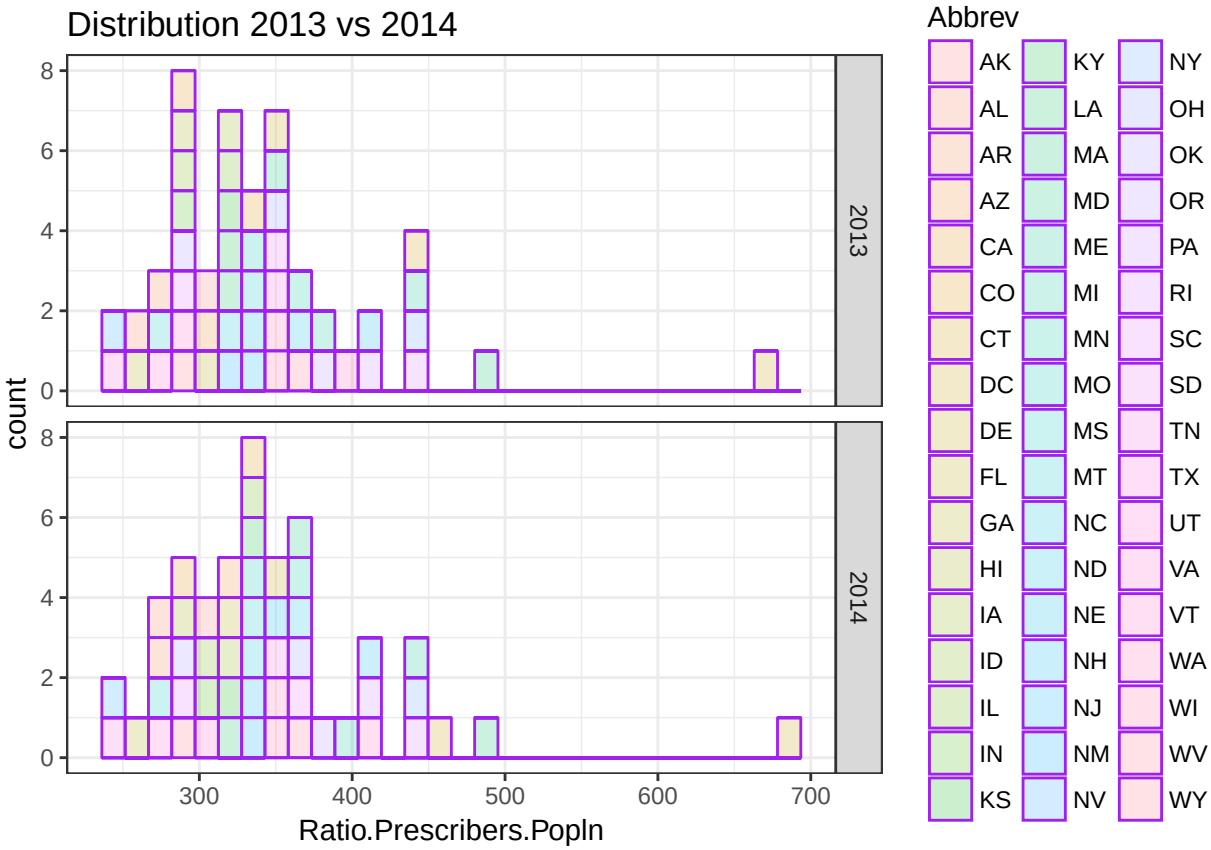

```
ggplot(Poverty.ODeaths.Prescribers.13.14, aes(x=as.numeric(`Opioid Prescribing Rate`))) + geom_histogra
```

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




```
ggplot(Poverty.ODeaths.Prescribers.13.14, aes(x=as.numeric(Ratio.Prescribers.Popln))) + geom_histogram
```

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




```
datax <- Poverty.ODeaths.Prescribers.13.14
```

```
OD$Population <- c(4833722,735132,6626624,2959373,38332521,5268367,3596080,925749,19552860,9992167,1404
OD$Deaths <- c(723, 124, 1211, 356, 4521, 899, 623, 189, 2634, 1206, 157, 212, 1705, 1172, 264, 332, 10
```

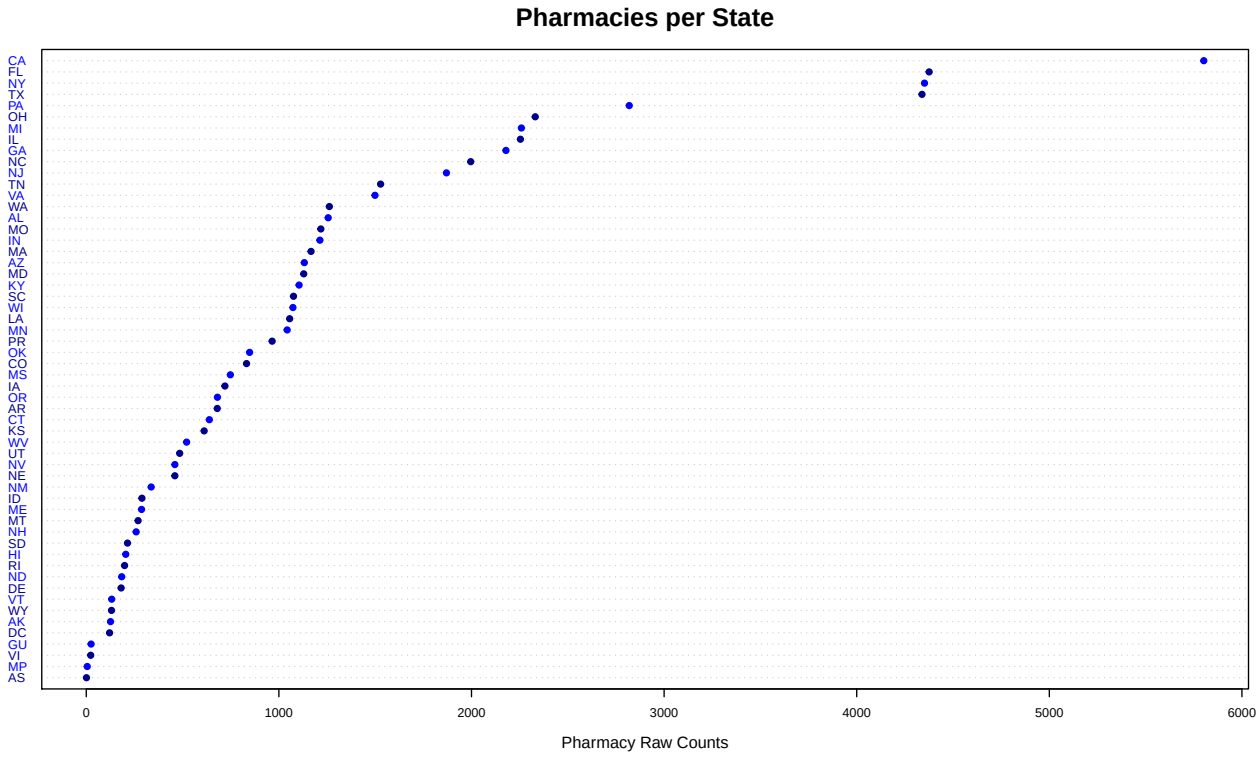
```
head(PharmacyStateCounts,2)
```

```
##    rn freq percentage
## 1: AK  126  0.2000858
## 2: AL 1256  1.9945056
```

```
PharmacyStateCounts$per100 <- round(100*as.numeric(PharmacyStateCounts$freq)/as.numeric(PharmacyStateCou
```

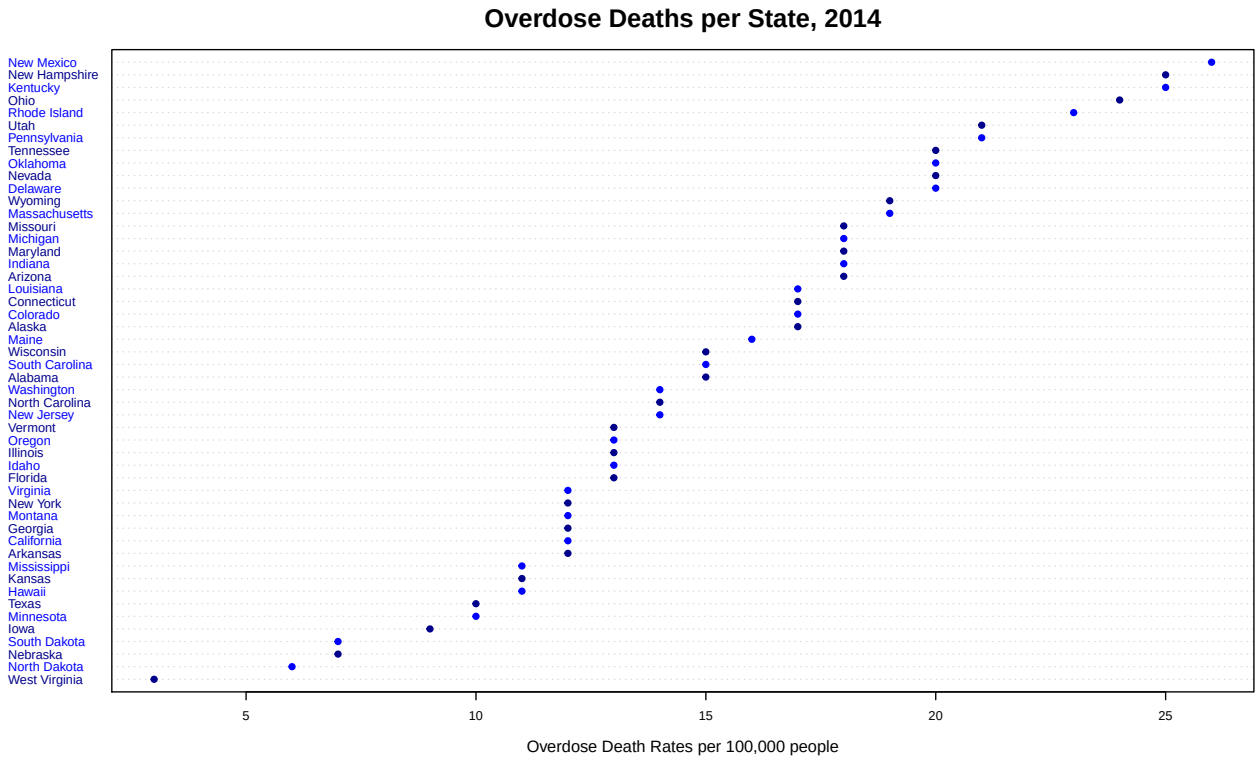
```
PharmacyStateCounts <- PharmacyStateCounts[order(PharmacyStateCounts$freq),]
```

```
dotchart(PharmacyStateCounts$freq, labels=PharmacyStateCounts$rn, cex=.6, main="Pharmacies per State",
xlab="Pharmacy Raw Counts", pch=19, col=c("darkblue","blue"), lcolor = "lightgrey",cex.main=2, cex.lab=
```




```
OD$DR <- round(100000*OD$Deaths/OD$Population)
OD <- OD[order(OD$DR),]
```

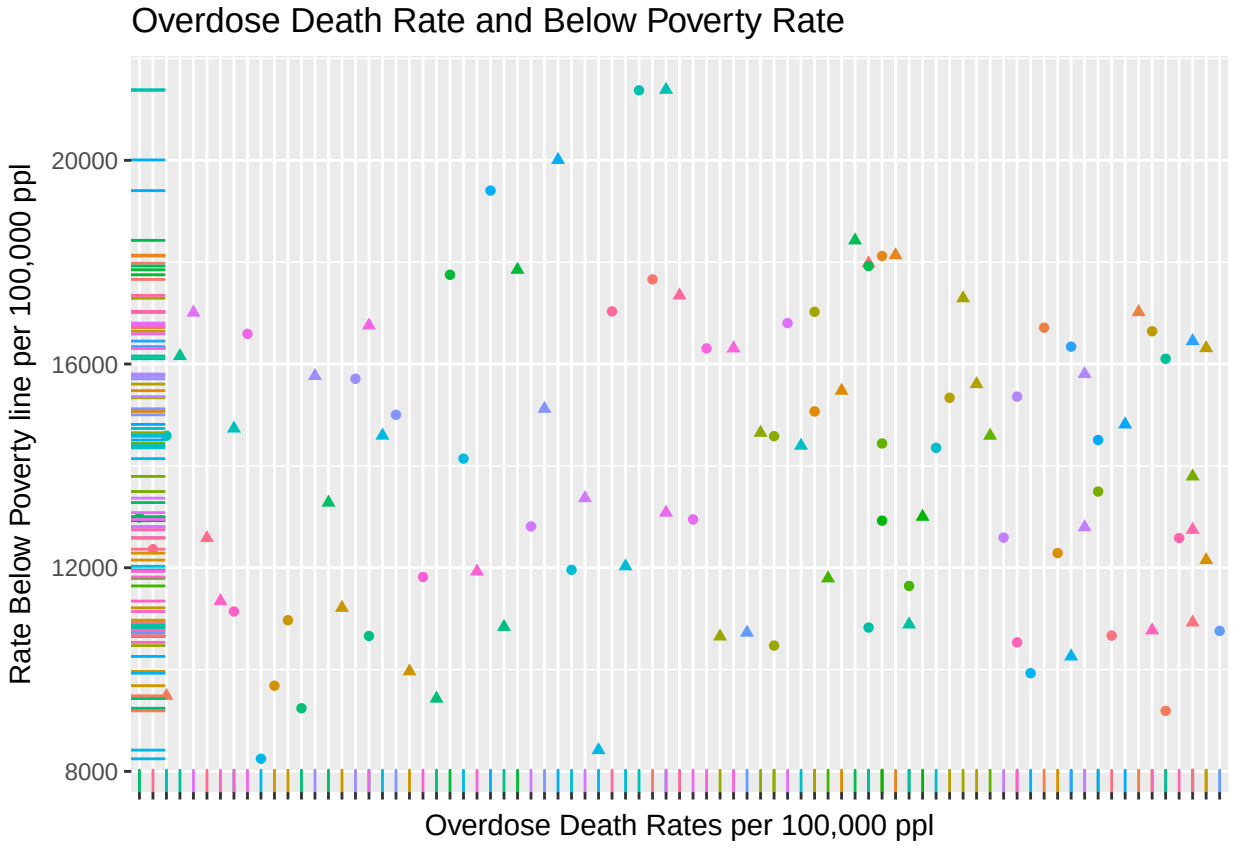
```
dotchart(OD$DR, labels=OD$State, cex=.6, main="Overdose Deaths per State, 2014",
xlab="Overdose Death Rates per 100,000 people", pch=19, col=c("darkblue","blue"), lcolor = "lightgrey",
```



```
Poverty.ODeaths.Prescribers.13.14$BelowpovRate <- round(100000*Poverty.ODeaths.Prescribers.13.14$`Below
Poverty.ODeaths.Prescribers.14 <- Poverty.ODeaths.Prescribers.13.14[Poverty.ODeaths.Prescribers.13.14$Y
```

```
# install.packages("car")
```

```
library(ggplot2)
ggplot(Poverty.ODeaths.Prescribers.13.14, aes(BelowpovRate, `Crude Rate Deaths`, col=State, shape=as.cha
```



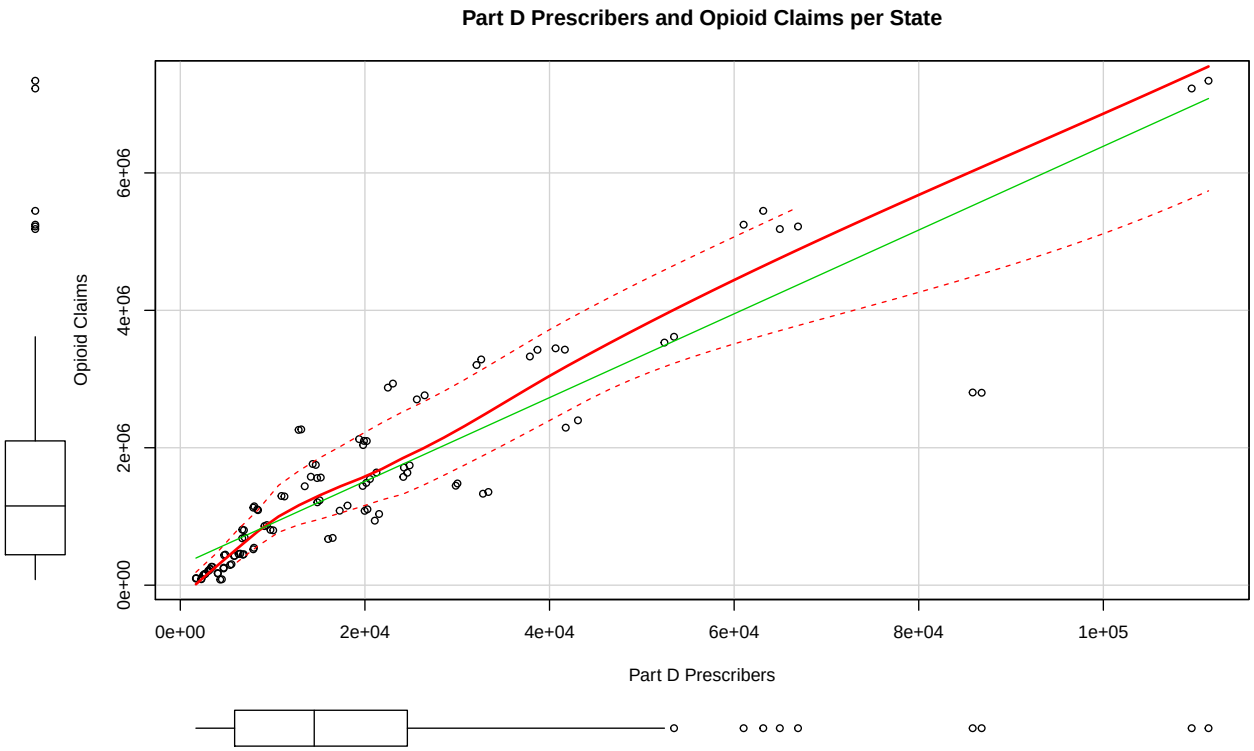

```
library(car)
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

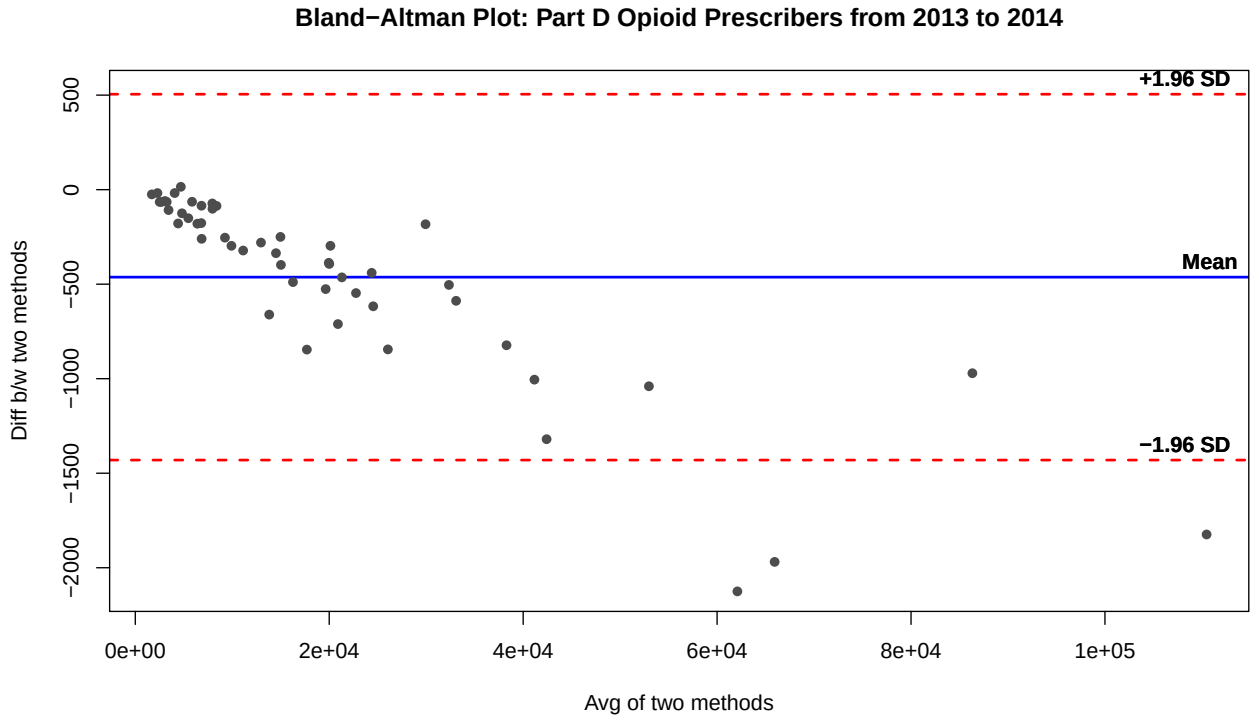
```
## The following object is masked from 'package:purrr':
##
##     some
```

```
attach(Poverty.ODeaths.Prescribers.13.14)
sp(`Part D Prescribers`, `Opioid Claims`, main = "Part D Prescribers and Opioid Claims per State")
```

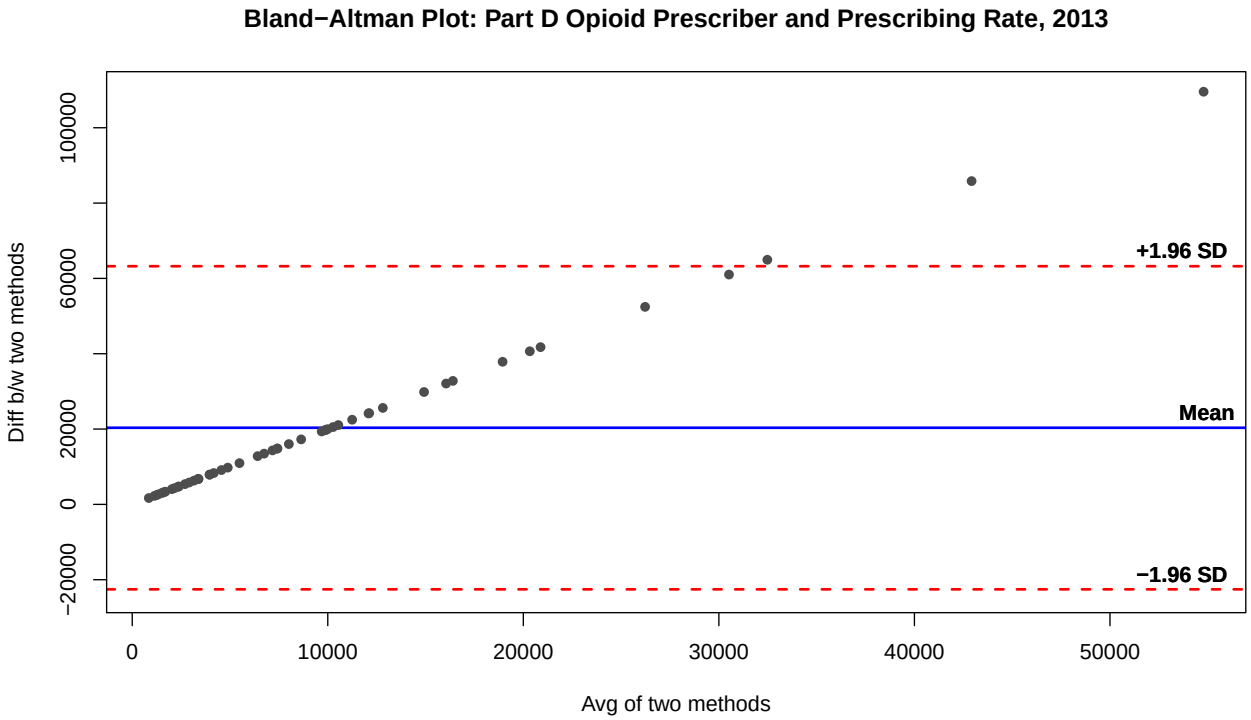



```
detach(Poverty.ODeaths.Prescribers.13.14)
```

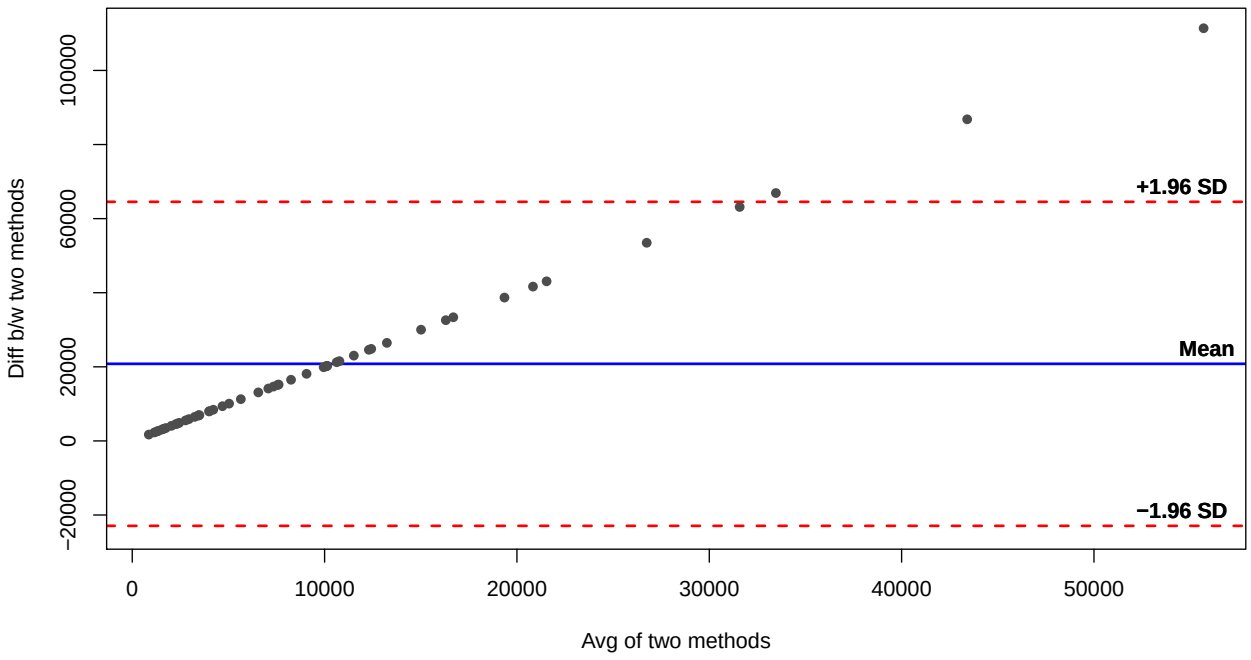
```
library(epade)
D_columns <- c("Part D Prescribers 2013","Opioid Prescribing Rate 2013","Opioid Prescribing Rate 2013",
# Small multiples, lines
bland.altman.ade(as.numeric(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014$`Part D Prescribers 20
```



```
bland.altman.ade(as.numeric(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014$`Part D Prescribers 2
```




```
bland.altman.ade(as.numeric(Part_D_Opioid_Prescribing_Change_Geographic_2013_2014$`Part D Prescribers 2
```



Do female or male doctors tend to prescribe more opioids?

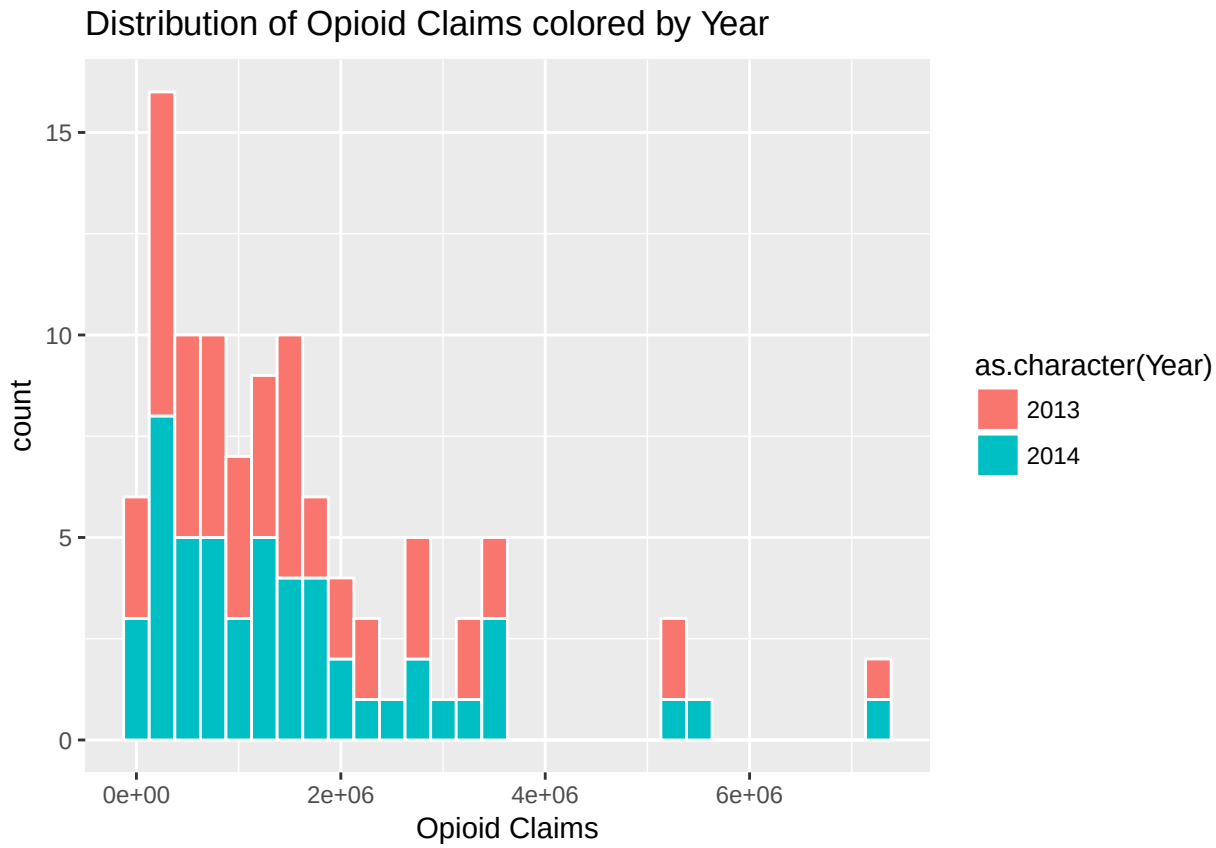
```
head(Poverty.ODeaths.Prescribers.13.14,3)
```

```
##     State Year Abbrev Part D Prescribers Opioid Claims Overall Claims
## 1 Alabama 2013     AL              12820       2260284       29160952
## 2 Alabama 2014     AL              13100       2267090       28852731
## 3  Alaska 2013     AK               2275         86517        1281057
##   Opioid Prescribing Rate Total Poverty
## 1                7.751064       4628774
## 2                7.857454       4644377
## 3                6.753564        693195
##   Below 100 percent of the poverty level
## 1                                 853751
## 2                                 871902
## 3                                  67553
##   100 to 149 percent of the poverty level
## 1                                  516964
## 2                                  519257
## 3                                   55420
##   At or above 150 percent of the poverty level Deaths Population
## 1                                      3258059    175    4833722
## 2                                      3253218    282    4849377
## 3                                       570222     69     735132
##   Crude Rate Deaths Crude Rate Lower 95% Confidence Interval
## 1               3.6                                     3.1
## 2               5.8                                     5.1
## 3               9.4                                     7.3
##   Crude Rate Upper 95% Confidence Interval
## 1                                      4.2
## 2                                      6.5
## 3                                     11.9
##   Prescriptions Dispensed by US Retailers in that year (millions)
## 1                                                             207
## 2                                                             196
## 3                                                             207
##   Opioid.Overall Ratio.Prescribers.Popln BelowpovRate
## 1     0.07751064                    265        17662
## 2     0.07857454                    270        17980
## 3     0.06753564                    309         9189
```

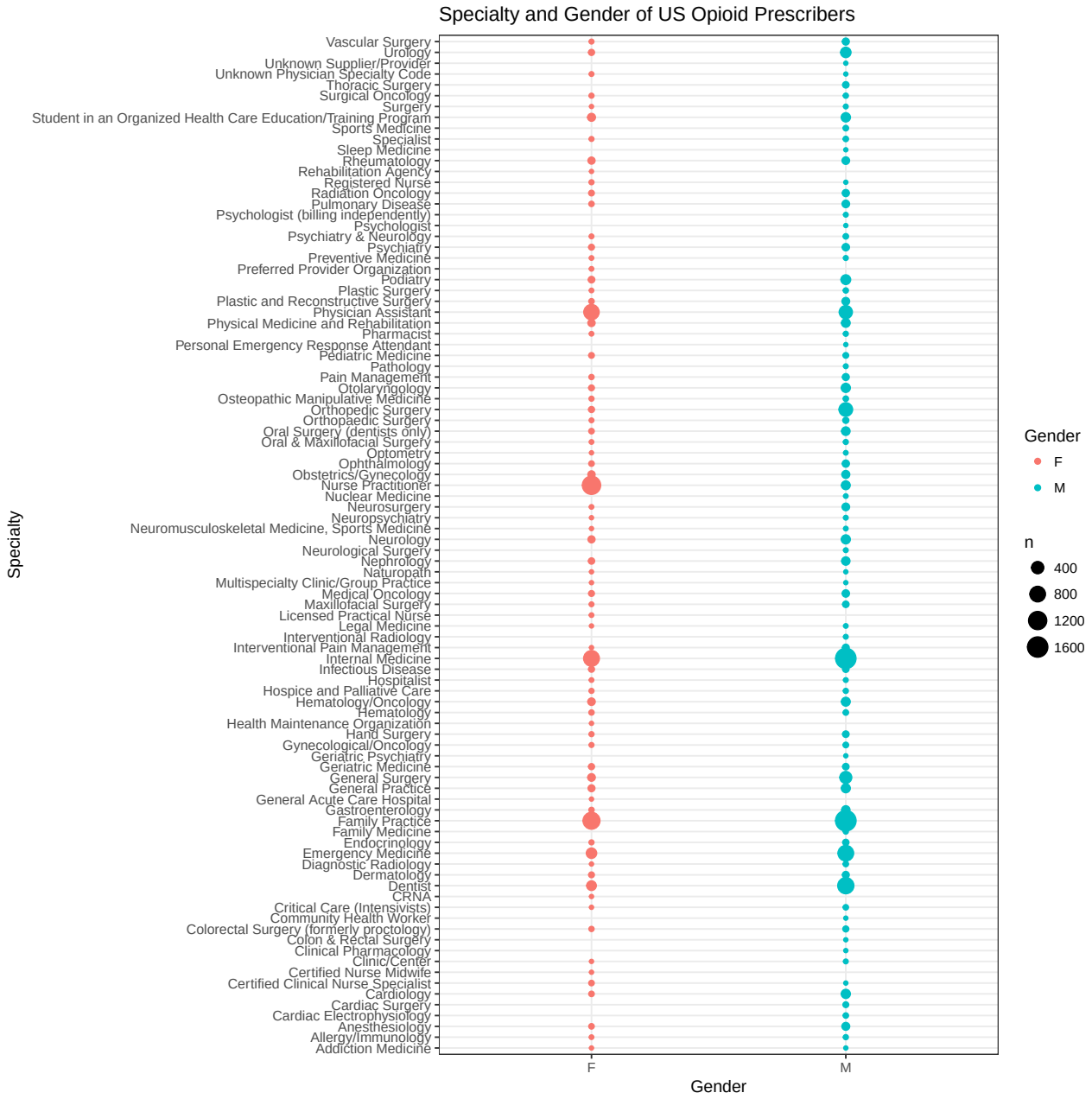
State Opioid Claims and Number of Deaths, colored by Year.

```
ggplot(Poverty.ODeaths.Prescribers.13.14, aes(x=`Opioid Claims`)) + geom_histogram(aes(fill=as.characte
```

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

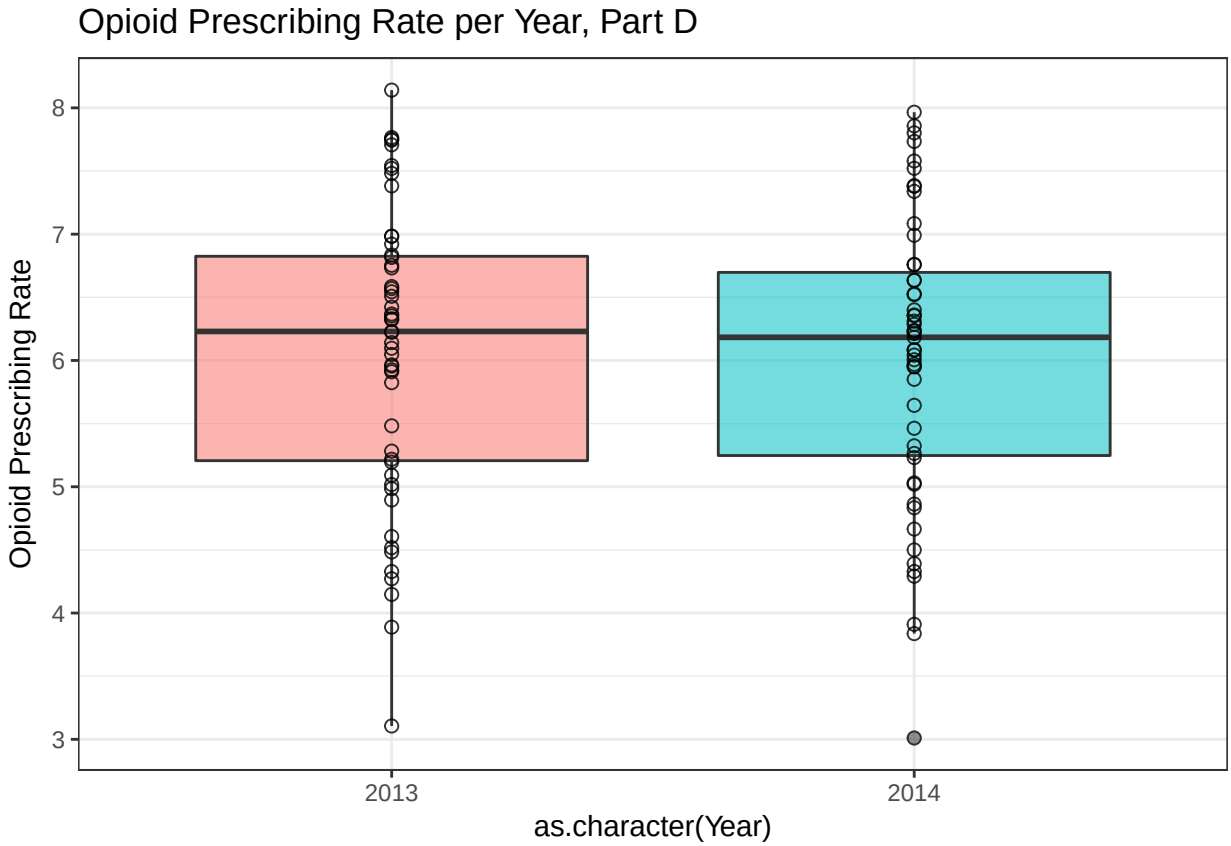



```
ggplot(Opioid.PrescribersDF, aes(Specialty,Gender, col=Gender)) + geom_count() + labs(title="Specialty a
```




```
df <- as.data.frame(multi.fun(prescriber_info$Specialty))
df <- df[order(df$freq),]
```

```
 # same plot as above but reordering by median values
  ggplot(data = Poverty.ODeaths.Prescribers.13.14, mapping = aes(x = as.character(Year), y = `Opioid Pr
    geom_boxplot(aes(fill=as.character(Year), alpha=0.2)) + geom_point(alpha=0.8, pch=21, size=2) + ther
```

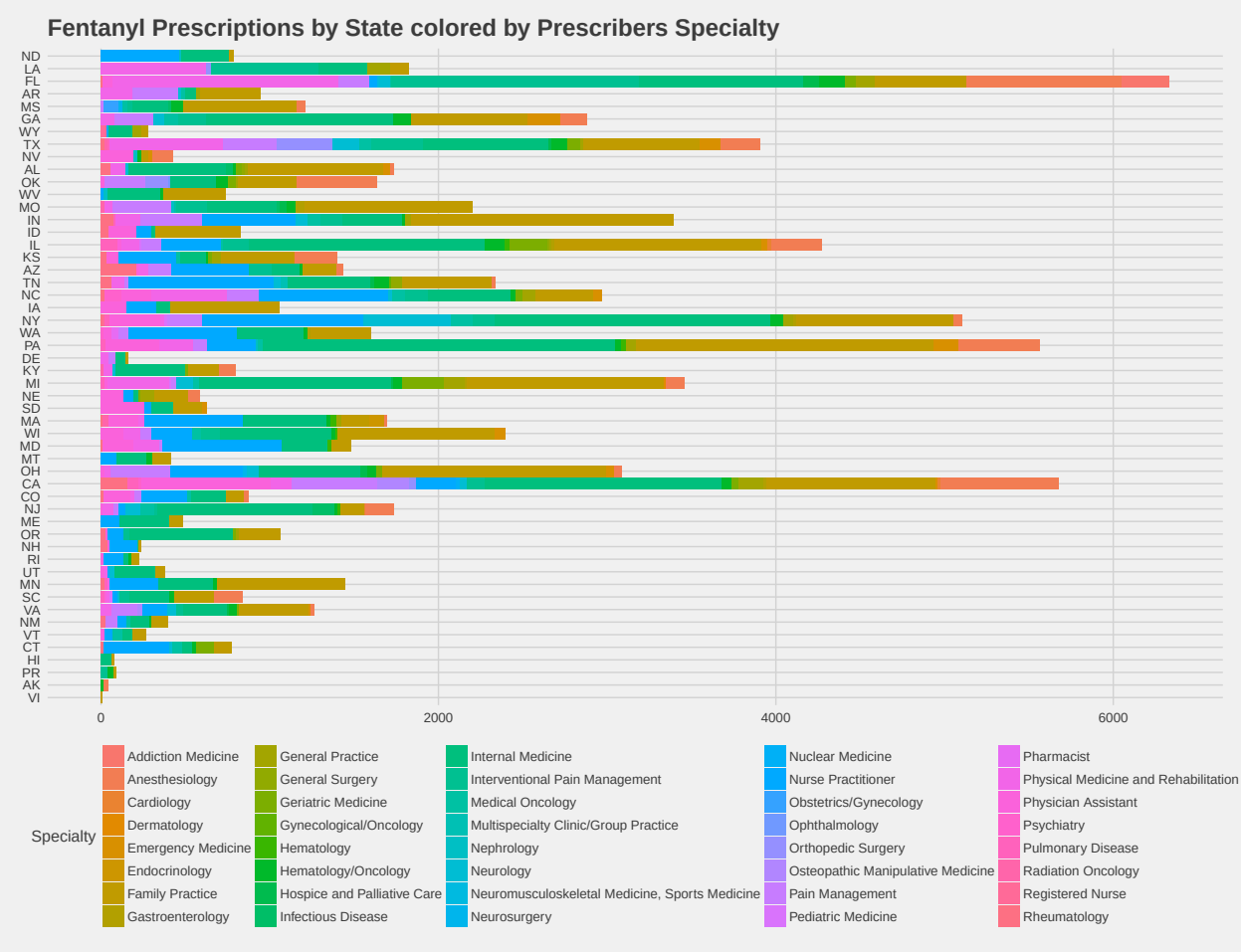



```
Fentanyl_Prescribers <- Opioid.PrescribersDF[Opioid.PrescribersDF$FENTANYL !=0,]
```

```
head(Fentanyl_Prescribers)
```

```
## # A tibble: 6 x 256
##          NPI Gender State   Credentials           Specialty ABILIFY
##        <int>  <chr> <chr>         <chr>               <chr>   <int>
## 1 1679650949      M    NV          M.D. Hematology/Oncology       0
## 2 1821106832      F    WI            MD   Internal Medicine       0
## 3 1144205303      M    CO            MD     Family Practice       0
## 4 1841349677      F    IN MSN, APRN, BC  Nurse Practitioner       0
## 5 1871682567      F    TN            NP  Nurse Practitioner      30
## 6 1629099304      M    SC            MD      Anesthesiology       0
## # ... with 250 more variables: ACETAMINOPHEN.CODEINE <int>,
## #   ACYCLOVIR <int>, ADVAIR.DISKUS <int>, AGGRENOX <int>,
## #   ALENDRONATE.SODIUM <int>, ALLOPURINOL <int>, ALPRAZOLAM <int>,
## #   AMIODARONE.HCL <int>, AMITRIPTYLINE.HCL <int>,
## #   AMLODIPINE.BESYLATE <int>, AMLODIPINE.BESYLATE.BENAZEPRIL <int>,
## #   AMOXICILLIN <int>, AMOX.TR.POTASSIUM.CLAVULANATE <int>,
## #   AMPHETAMINE.SALT.COMBO <int>, ATENOLOL <int>,
## #   ATORVASTATIN.CALCIUM <int>, AVODART <int>, AZITHROMYCIN <int>,
## #   BACLOFEN <int>, BD.ULTRA.FINE.PEN.NEEDLE <int>, BENAZEPRIL.HCL <int>,
## #   BENICAR <int>, BENICAR.HCT <int>, BENZTROPINE.MESYLATE <int>,
## #   BISOPROLOL.HYDROCHLOROTHIAZIDE <int>, BRIMONIDINE.TARTRATE <int>,
## #   BUMETANIDE <int>, BUPROPION.HCL.SR <int>, BUPROPION.XL <int>,
## #   BUSPIRONE.HCL <int>, BYSTOLIC <int>, CARBAMAZEPINE <int>,
## #   CARBIDOPA.LEVODOPA <int>, CARISOPRODOL <int>, CARTIA.XT <int>,
```

```
## #   CARVEDILOL <int>, CEFUROXIME <int>, CELEBREX <int>, CEPHALEXIN <int>,
## #   CHLORHEXIDINE.GLUCONATE <int>, CHLORTHALIDONE <int>, CILOSTAZOL <int>,
## #   CIPROFLOXACIN.HCL <int>, CITALOPRAM.HBR <int>, CLINDAMYCIN.HCL <int>,
## #   CLOBETASOL.PROPIONATE <int>, CLONAZEPAM <int>, CLONIDINE.HCL <int>,
## #   CLOPIDOGREL <int>, CLOTRIMAZOLE.BETAMETHASONE <int>, COLCRYS <int>,
## #   COMBIVENT.RESPIMAT <int>, CRESTOR <int>, CYCLOBENZAPRINE.HCL <int>,
## #   DEXILANT <int>, DIAZEPAM <int>, DICLOFENAC.SODIUM <int>,
## #   DICYCLOMINE.HCL <int>, DIGOX <int>, DIGOXIN <int>,
## #   DILTIAZEM.24HR.CD <int>, DILTIAZEM.24HR.ER <int>, DILTIAZEM.ER <int>,
## #   DILTIAZEM.HCL <int>, DIOVAN <int>, DIPHENOXYLATE.ATROPINE <int>,
## #   DIVALPROEX.SODIUM <int>, DIVALPROEX.SODIUM.ER <int>,
## #   DONEPEZIL.HCL <int>, DORZOLAMIDE.TIMOLOL <int>,
## #   DOXAZOSIN.MESYLATE <int>, DOXEPIN.HCL <int>,
## #   DOXYCYCLINE.HYCLATE <int>, DULOXETINE.HCL <int>,
## #   ENALAPRIL.MALEATE <int>, ESCITALOPRAM.OXALATE <int>, ESTRADIOL <int>,
## #   EXELON <int>, FAMOTIDINE <int>, FELODIPINE.ER <int>,
## #   FENOFIBRATE <int>, FENTANYL <int>, FINASTERIDE <int>,
## #   FLOVENT.HFA <int>, FLUCONAZOLE <int>, FLUOXETINE.HCL <int>,
## #   FLUTICASONE.PROPIONATE <int>, FUROSEMIDE <int>, GABAPENTIN <int>,
## #   GEMFIBROZIL <int>, GLIMEPIRIDE <int>, GLIPIZIDE <int>,
## #   GLIPIZIDE.ER <int>, GLIPIZIDE.XL <int>, GLYBURIDE <int>,
## #   HALOPERIDOL <int>, HUMALOG <int>, HYDRALAZINE.HCL <int>,
## #   HYDROCHLOROTHIAZIDE <int>, HYDROCODONE.ACETAMINOPHEN <int>, ...
ggplot(Fentanyl_Prescribers, aes(reorder(State, FENTANYL), FENTANYL)) + geom_col(aes(fill=Specialty)) +
```




```
# not.include <- c("")
# df_mat <- as.matrix(Opioid.PrescribersDF)
# cors <- cor(df_mat[,6:13])
# col<- colorRampPalette(c('blue', 'white', 'red'))(20)
# heatmap(x = cors, col = col, symm = T)
```

Where the data is from:

http://wonder.cdc.gov/wonder

Citation for Opioid Prescription Data: IMS Health, Vector One: National, years 1991-1996, Data Extracted 2011. IMS Health, National Prescription Audit, years 1997-2013, Data Extracted 2014. Accessed at NIDA article linked (Figure 1) on Oct 23, 2016.

Opioid Prescriptions Dispensed by US Retail Pharmacies 1991-2013

This data includes the number of opioid prescriptions dispensed (millions) by US retail pharmacies from 1991-2013. The figures were taken from the diagram above in an article by Nora D. Volkow, M.D. on the National Institute of Drug Abuse.

US retail pharmacies, Q42009-Q22015

This data includes the number of opioid analgesic prescriptions dispensed by US retail pharmacies from Q4 2009 to Q2 2015. This dataset includes breakdowns by type of opioid analgesic. (Added 26 Oct 2016)

Demographic Data comes from the Census using the query tool 'FactFinder' to search for datasets.