Reproduced analysis of FARS data

Marialexia Alfano

November 4, 2017

library(tidyr)

## Warning: package 'tidyr' was built under R version 3.3.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.3.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.3

library(ggthemes)

## Warning: package 'ggthemes' was built under R version 3.3.3

library(knitr)

## Warning: package 'knitr' was built under R version 3.3.3

library(purrr)

## Warning: package 'purrr' was built under R version 3.3.3

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.3.3

## Loading tidyverse: tibble  
## Loading tidyverse: readr

## Warning: package 'tibble' was built under R version 3.3.3

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

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

library(tibble)  
  
load("../data/clean\_fars.RData")  
source("../R/fars\_functions.R")

## Results

### Percentages of drivers testing positive by drug type, sex, and year group

#table 1   
clean\_fars %>%   
 mutate(year\_cat = cut(year, breaks = c(1999, 2002, 2006, 2010),  
 labels = c("1999-2002","2003-2006",  
 "2007-2010"),  
 include.lowest = TRUE, right = TRUE)) %>%   
 filter(!is.na(sex)) %>%   
 group\_by(drug\_type, sex, year\_cat) %>%   
 summarize(n\_non\_missing = sum(!is.na(positive\_for\_drug)),  
 positive\_test = sum(positive\_for\_drug, na.rm = TRUE),  
 perc\_positive = round(100 \* positive\_test/n\_non\_missing, 1)) %>%   
 select(drug\_type, sex, year\_cat, perc\_positive) %>%   
 unite(sex\_year\_cat, sex, year\_cat) %>%   
 spread(sex\_year\_cat, perc\_positive) %>%   
 knitr::kable(col.names = c("Drug Type", "F 1999-2002",  
 "F 2003-2006", "F 2007-2010",  
 "M 1999-2002", "M 2003-2006",  
 "M 2007-2010"))

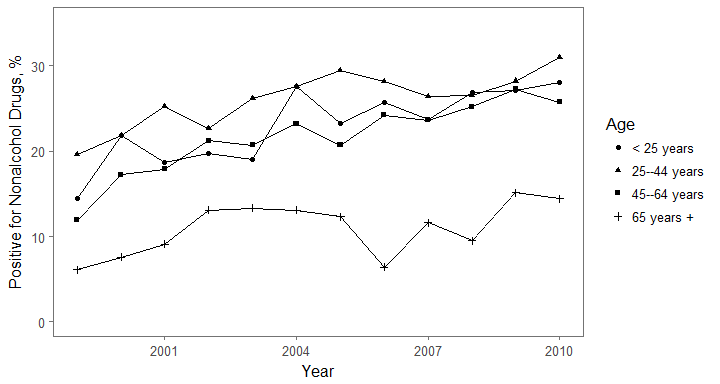
## Warning: package 'bindrcpp' was built under R version 3.3.3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Drug Type | F 1999-2002 | F 2003-2006 | F 2007-2010 | M 1999-2002 | M 2003-2006 | M 2007-2010 |
| Alcohol | 26.4 | 24.3 | 27.1 | 43.2 | 42.9 | 43.3 |
| Cannabinoid | 2.8 | 5.7 | 7.3 | 5.8 | 10.3 | 11.8 |
| Depressant | 3.4 | 3.8 | 4.8 | 2.0 | 2.5 | 3.2 |
| Narcotic | 4.2 | 4.9 | 7.0 | 2.2 | 3.4 | 4.0 |
| Other | 5.6 | 6.6 | 7.2 | 4.3 | 4.5 | 4.2 |
| Stimulant | 7.2 | 9.1 | 8.7 | 10.5 | 11.9 | 9.2 |

### Figure 1: Prevalence of nonalcohol drugs in fatally injured drivers by year and age group

#adjusting data for figure one   
figure\_one <- clean\_fars %>%   
 filter(drug\_type != "Alcohol") %>%  
 filter(!is.na(agecat)) %>%  
 group\_by(unique\_id, year, agecat) %>%   
 summarize(positive\_tests = sum(positive\_for\_drug, na.rm = TRUE), positive = any(positive\_tests > 0)) %>%   
 ungroup() %>%   
 group\_by(year, agecat) %>%   
 summarize(total\_tests = length(positive), positive = sum(positive), percent\_positive = round(100 \* positive/total\_tests,1))

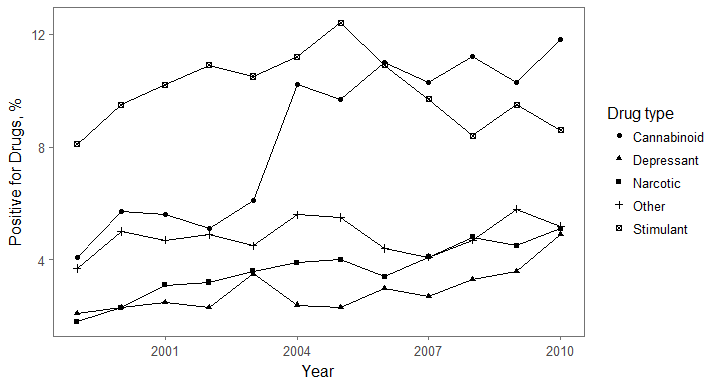
#plot for figure one  
plot\_figure\_one <- figure\_one %>%  
 ggplot(aes(x = year, y = percent\_positive, position = agecat))+  
 geom\_point(aes(shape = agecat))+  
 geom\_line()+  
 theme\_few()+  
 labs(x = "Year", y = "Positive for Nonalcohol Drugs, %", shape = "Age")+  
 scale\_y\_continuous(limits = c(0, 35))  
   
plot\_figure\_one



### Figure 2: Prevalence of nonalcohol drugs in fatally injured drivers by year and drug type

#adjusting data for figure two  
figure\_two <- clean\_fars %>%   
 filter(drug\_type != "Alcohol") %>%   
 filter(!is.na(positive\_for\_drug)) %>%   
 group\_by(drug\_type, year) %>%   
 summarize(n\_not\_NA = sum(!is.na(positive\_for\_drug)),  
 positive\_tests = sum(positive\_for\_drug, na.rm = TRUE),  
 percent\_positive = round(100 \* positive\_tests/n\_not\_NA,1))

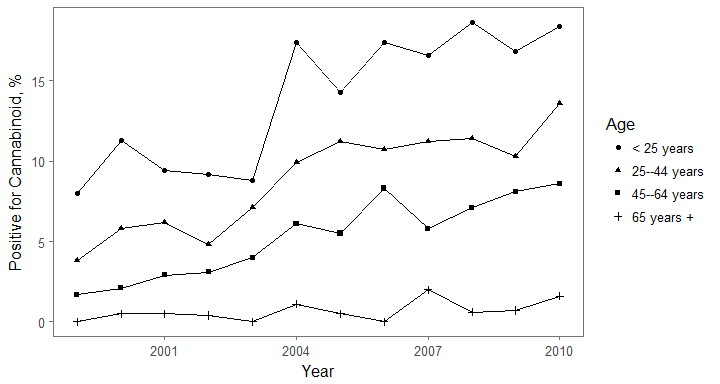
#plot for figure two  
plot\_figure\_two <- figure\_two %>%   
 ggplot(aes(x = year, y = percent\_positive, position = drug\_type))+  
 geom\_point(aes(shape = drug\_type))+  
 geom\_line()+  
 theme\_few()+  
 scale\_y\_continuous(breaks = c(0, 4, 8, 12))+  
 labs(x = "Year", y = "Positive for Drugs, %", shape = "Drug type")  
  
  
plot\_figure\_two



### Figure 3: Prevalence of cannabinoid drugs in fatally injured drivers by year and age group

#adjusting data for figure three  
figure\_three <- clean\_fars %>%  
 filter(drug\_type == "Cannabinoid") %>%   
 filter(!is.na(agecat)) %>%   
 group\_by(agecat, year) %>%   
 summarize(n\_not\_NA = sum(!is.na(positive\_for\_drug)),  
 positive\_tests = sum(positive\_for\_drug, na.rm = TRUE),  
 percent\_positive = round(100 \* positive\_tests/n\_not\_NA,1))

#plot for figure three  
plot\_figure\_three <- figure\_three %>%   
 ggplot(aes(x = year, y = percent\_positive, position = agecat))+  
 geom\_line()+  
 geom\_point(aes(shape = agecat))+  
 theme\_few()+  
 labs(x = "Year", y = "Positive for Cannabinoid, %", shape = "Age")  
  
plot\_figure\_three



### Prevalence of drugs in fatally injured drivers for 1999 and 2010 by drug type

#table 2  
prev\_table\_one <- clean\_fars %>%   
 filter(year %in% c("1999","2010")) %>%   
 group\_by(year, drug\_type) %>%   
 summarize(positive = sum(positive\_for\_drug, na.rm = TRUE), trials = sum(!is.na(positive\_for\_drug)))  
  
prev\_table\_two <- data.frame(prevalence = perc\_cis(x = prev\_table\_one$positive, n = prev\_table\_one$trials))  
  
prev\_table\_one <- rowid\_to\_column(as.data.frame(prev\_table\_one, var = "rowid"))  
prev\_table\_two <- rowid\_to\_column(as.data.frame(prev\_table\_two, var = "rowid"))  
  
prev\_table <- prev\_table\_one %>%   
 full\_join(prev\_table\_two, by = "rowid") %>%   
 select(prevalence, drug\_type, year) %>%   
 spread(key = year, value = prevalence) %>%   
 rename("Drug type" = drug\_type)  
  
kable(prev\_table)

|  |  |  |
| --- | --- | --- |
| Drug type | 1999 | 2010 |
| Alcohol | 38.7% (36.5%, 40.9%) | 39.1% (36.7%, 41.5%) |
| Cannabinoid | 4.1% (3.1%, 5%) | 11.8% (10.2%, 13.4%) |
| Depressant | 2.1% (1.5%, 2.8%) | 4.9% (3.8%, 5.9%) |
| Narcotic | 1.8% (1.2%, 2.5%) | 5.1% (4%, 6.2%) |
| Other | 3.7% (2.9%, 4.6%) | 5.2% (4.1%, 6.3%) |
| Stimulant | 8.1% (6.8%, 9.4%) | 8.6% (7.2%, 9.9%) |

## Statistics for testing for trend in prevalence of drugs over study years by drug type using Cochran-Armitage trend test

#table 3  
drug\_list <- c("Alcohol", "Nonalcohol", "Narcotic", "Depressant",  
 "Stimulant", "Cannabinoid", "Other")  
drug\_trend\_tests\_ca <- lapply(drug\_list, test\_trend\_ca)  
drug\_trend\_tests\_ca <- dplyr::bind\_rows(drug\_trend\_tests\_ca) %>%  
 dplyr::mutate(drug = drug\_list) %>%  
 dplyr::select(drug, Z, p.value)  
drug\_trend\_tests\_ca %>% knitr::kable()

|  |  |  |
| --- | --- | --- |
| drug | Z | p.value |
| Alcohol | 1.2 | 0.228 |
| Nonalcohol | 9.7 | 0.000 |
| Narcotic | 6.7 | 0.000 |
| Depressant | 4.7 | 0.000 |
| Stimulant | 0.5 | 0.604 |
| Cannabinoid | 13.6 | 0.000 |
| Other | 1.4 | 0.157 |

### Statistics for testing for trend in prevalence of drugs over study years by drug type using Wald test of logistic regression coefficient for "year"

#table 4  
drug\_list <- c("Alcohol", "Nonalcohol", "Narcotic", "Depressant",  
"Stimulant", "Cannabinoid", "Other")  
drug\_trend\_tests\_log\_reg <- lapply(drug\_list, test\_trend\_log\_reg)  
drug\_trend\_tests\_log\_reg <- dplyr::bind\_rows(drug\_trend\_tests\_log\_reg) %>%  
dplyr::mutate(drug = drug\_list) %>%  
dplyr::select(drug, Z, p.value)  
drug\_trend\_tests\_log\_reg %>% knitr::kable()

|  |  |  |
| --- | --- | --- |
| drug | Z | p.value |
| Alcohol | 1.2 | 0.228 |
| Nonalcohol | 9.9 | 0.000 |
| Narcotic | 6.6 | 0.000 |
| Depressant | 4.7 | 0.000 |
| Stimulant | -0.5 | 0.604 |
| Cannabinoid | 13.5 | 0.000 |
| Other | 1.4 | 0.158 |