Reproduced analysis of FARS data

Kate Huebner, Maggie Weinroth, Colleen Brents

10/27/2017

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE,
                      warning = FALSE, error = FALSE)
```

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(ggthemes)
```

```
load("../data/clean_fars.RData")
source("../R/fars_functions.R")
dim(clean_fars)
```

```
## [1] 156413      6
```

```
length(unique(clean_fars$unique_id))
```

```
## [1] 25593
```

```
summary(clean_fars)
```

```
##   unique_id             sex              year               agecat
##  Length:156413      Male  :121072   Min.   :1999   < 25 years  :39149
##  Class :character   Female: 35335   1st Qu.:2002   25--44 years:61235
##  Mode  :character   NA's  :     6   Median :2004   45--64 years:39870
##                                     Mean   :2004   65 years +  :16108
##                                     3rd Qu.:2007   NA's        :   51
##                                     Max.   :2010
##        drug_type     positive_for_drug
##  Alcohol    :25593   Mode :logical
##  Cannabinoid:26260   FALSE:127597
##  Depressant :25988   TRUE :16894
##  Narcotic   :26086   NA's :11922
##  Other      :26179
##  Stimulant  :26307
```

```
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
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 1
                           "M 2007-2010"))
```

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

Drug type	F 1999-2002	F 2003-2006	F 2007-2010	M 1999-2002	M 2003-2006	M 2007-2010
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

```
nonalcohol_data <- clean_fars %>%
filter(!drug_type == "Alcohol") %>%
select(unique_id, drug_type, year, agecat, positive_for_drug) %>%
filter(!is.na(positive_for_drug)) %>%
filter(!is.na(agecat)) %>%
group_by(year, agecat, unique_id) %>%
summarize(positive_for_drug = any(positive_for_drug)) %>%
ungroup() %>%
group_by(year, agecat) %>%
summarize(count = 100*(mean(positive_for_drug)))

nonalcohol_data
```

```
## # A tibble: 48 x 3
## # Groups:   year [?]
##     year       agecat     count
##    <int>       <fctr>     <dbl>
##  1  1999   < 25 years 15.618221
##  2  1999 25--44 years 21.592443
##  3  1999 45--64 years 13.725490
##  4  1999   65 years +  7.425743
##  5  2000   < 25 years 23.908046
##  6  2000 25--44 years 24.258760
##  7  2000 45--64 years 19.585253
##  8  2000   65 years +  9.183673
##  9  2001   < 25 years 20.612245
## 10  2001 25--44 years 27.706186
## # ... with 38 more rows
```

```
plot_1 <- nonalcohol_data %>%
ggplot(aes(x = year, y = count, shape = agecat)) +
geom_line(linetype=1) +
geom_point(aes(shape=agecat), size=2, col="black") +
scale_shape(solid=TRUE) +
labs(x = "Year", y = "Positive for Nonalcohol Drugs, %)", shape = "Age") +
#scale_y_continuous(limits = c(0,40), breaks = 10, name = "") +
theme_few()
plot_1
```

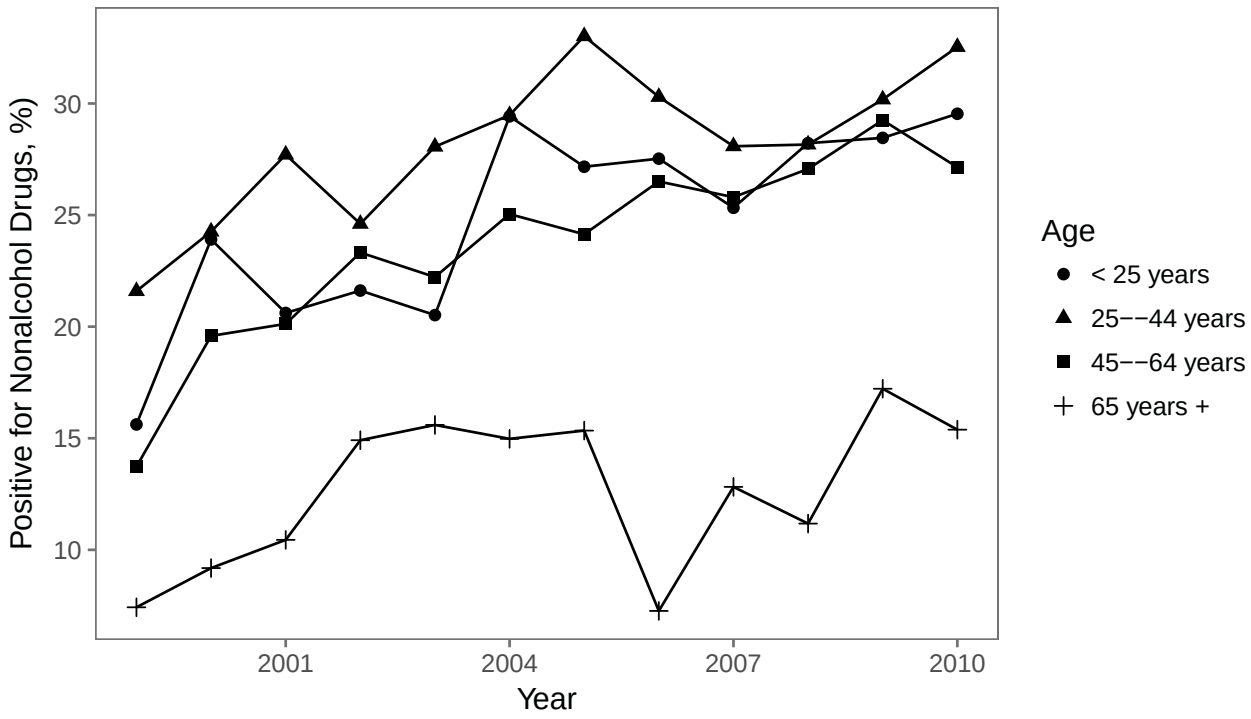

Figure 2: Prevalence of nonalcohol drugs in fatally injured drivers by year and age group

```
data2 <- clean_fars %>%
filter(!drug_type == "Alcohol") %>%
select(unique_id, drug_type, year, agecat, positive_for_drug) %>%
filter(!is.na(positive_for_drug)) %>%
filter(!is.na(agecat)) %>%
group_by(year, drug_type, unique_id) %>%
summarize(positive_for_drug = any(positive_for_drug)) %>%
ungroup() %>%
group_by(year, drug_type) %>%
summarize(count = 100*(mean(positive_for_drug)))

nonalcohol_data
```

```
## # A tibble: 48 x 3
## # Groups:   year [?]
##     year       agecat     count
##    <int>       <fctr>     <dbl>
##  1  1999   < 25 years 15.618221
##  2  1999 25--44 years 21.592443
##  3  1999 45--64 years 13.725490
##  4  1999   65 years +  7.425743
##  5  2000   < 25 years 23.908046
##  6  2000 25--44 years 24.258760
##  7  2000 45--64 years 19.585253
##  8  2000   65 years +  9.183673
##  9  2001   < 25 years 20.612245
## 10  2001 25--44 years 27.706186
## # ... with 38 more rows
```

```
plot_2 <- data2 %>%
ggplot(aes(x = year, y = count, shape = drug_type)) +
geom_line(linetype=1) +
geom_point(aes(shape= drug_type), size=2, col="black") +
scale_shape(solid=TRUE) +
labs(x = "Year", y = "Positive for Drugs, %)", shape = "Drug type") +
theme_few()
plot_2
```

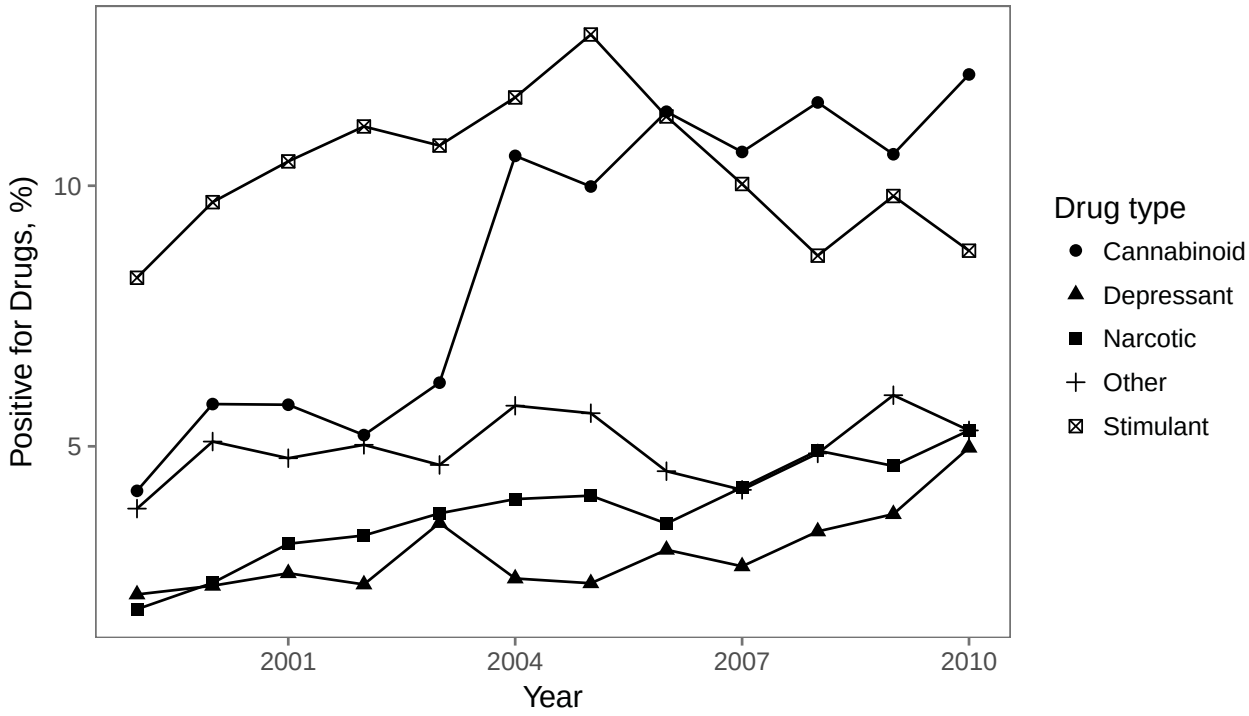

Figure 3: Prevalence of cannabinoid drugs in fatally injured drivers by year and age group

```
plot_data <- clean_fars %>%
filter(drug_type == "Cannabinoid") %>%
select(unique_id, drug_type, year, agecat, positive_for_drug) %>%
filter(!is.na(positive_for_drug)) %>%
filter(!is.na(agecat)) %>%
group_by(year, agecat, unique_id) %>%
summarize(positive_for_drug = any(positive_for_drug)) %>%
ungroup() %>%
group_by(year, agecat) %>%
summarize(count = 100*(mean(positive_for_drug)))

plot_data
```

```
## # A tibble: 48 x 3
## # Groups:   year [?]
##     year       agecat      count
##    <int>       <fctr>      <dbl>
##  1  1999   < 25 years  8.2429501
##  2  1999 25--44 years  3.9136302
##  3  1999 45--64 years  1.6806723
```

```
##  4  1999   65 years +  0.0000000
##  5  2000   < 25 years 11.7241379
##  6  2000 25--44 years  5.9299191
##  7  2000 45--64 years  2.0737327
##  8  2000   65 years +  0.5102041
##  9  2001   < 25 years  9.7959184
## 10  2001 25--44 years  6.4432990
## # ... with 38 more rows
```

```
plot_3 <- plot_data %>%
ggplot(aes(x = year, y = count, shape = agecat)) +
geom_line(linetype=1) +
geom_point(aes(shape=agecat), size=2, col="black") +
scale_shape(solid=TRUE) +
labs(x = "Year", y = "Positive for Cannabiniod, %)", shape = "Age") +
theme_few()
plot_3
```

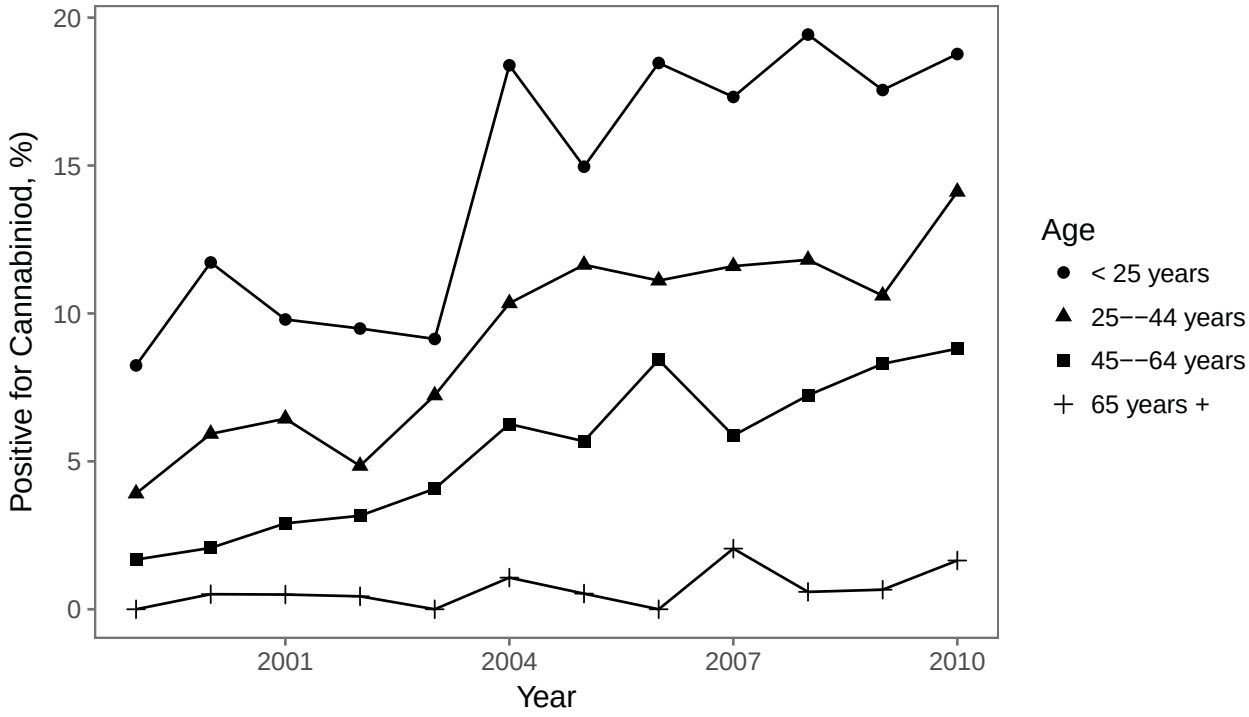

Figure 4: Prevalence of drugs in fatally injured drivers for 1999 and 2010 by drug type

```
#first function table
table1 <- clean_fars %>%
  filter(year %in% c(1999, 2010)) %>%
group_by(drug_type, year) %>%
  summarize(n_non_missing = sum(!is.na(positive_for_drug)),
positive_test = sum(positive_for_drug, na.rm = TRUE),
percent = perc_cis(x = positive_test, n = n_non_missing)) %>%
  select(drug_type,year, percent) %>%
#unite(year, percent) %>%
spread(key = year, value = percent) %>%
knitr::kable(col.names = c("Drug type", "1999", "2010"))
```

```
table1
```

Drug type	1999	2010
Alcohol	38.7% (37.8%, 39.5%)	39.1% (38.1%, 40%)
Cannabinoid	4.1% (4%, 4.1%)	11.8% (11.6%, 12%)
Depressant	2.1% (2.1%, 2.1%)	4.9% (4.8%, 4.9%)
Narcotic	1.8% (1.8%, 1.9%)	5.1% (5.1%, 5.2%)
Other	3.7% (3.7%, 3.8%)	5.2% (5.1%, 5.2%)
Stimulant	8.1% (8%, 8.2%)	8.6% (8.4%, 8.7%)

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

```
#second function table
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(digits = 2)
```

drug	Z	p.value
Alcohol	1.21	0.23
Nonalcohol	11.06	0.00
Narcotic	6.66	0.00
Depressant	4.72	0.00
Stimulant	0.52	0.60
Cannabinoid	13.59	0.00
Other	1.41	0.16

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

```
#Third function table
drug_list <- c("Alcohol", "Nonalcohol", "Narcotic", "Depressant",
               "Stimulant", "Cannabinoid", "Other")
drug_trend_tests_ca <- lapply(drug_list, test_trend_log_reg)
drug_trend_tests_ca <- dplyr::bind_rows(drug_trend_tests_ca) %>%
  dplyr::mutate(drug = drug_list) %>%
  dplyr::select(drug, "z value", "Pr(>|z|)")
drug_trend_tests_ca %>%
  knitr::kable(digits = 2)
```

drug	z value	Pr(>|z|)
Alcohol	1.21	0.23
Nonalcohol	11.03	0.00
Narcotic	6.63	0.00
Depressant	4.70	0.00
Stimulant	-0.52	0.60

drug	z value	Pr(>|z|)
Cannabinoid	13.47	0.00
Other	1.41	0.16