Vegnette for fars package for Coursera

This vegnette is used for the `fars` package for the assignment of the course "r packages" on Coursera. The main R file is given.

The main usage of this package is to plot the accident rate on the map for each State in United States.

The sturcture of data is like this

```
library(dplyr)
library(readr)
dt <- read_csv("accident_2013.csv.bz2")
dt
```

```
## # A tibble: 30,202 x 50
##    STATE ST_CASE VE_TOTAL VE_FORMS PVH_INVL  PEDS PERNOTMVIT PERMVIT
##    <int>   <int>    <int>    <int>    <int> <int>      <int>   <int>
##  1     1   10001        1        1        0     0          0       8
##  2     1   10002        2        2        0     0          0       2
##  3     1   10003        1        1        0     0          0       1
##  4     1   10004        1        1        0     0          0       3
##  5     1   10005        2        2        0     0          0       3
##  6     1   10006        2        2        0     0          0       3
##  7     1   10007        1        1        0     0          0       1
##  8     1   10008        2        2        0     0          0       2
##  9     1   10009        1        1        0     0          0       1
## 10     1   10010        2        2        0     0          0       4
## # ... with 30,192 more rows, and 42 more variables: PERSONS <int>,
## #   COUNTY <int>, CITY <int>, DAY <int>, MONTH <int>, YEAR <int>,
## #   DAY_WEEK <int>, HOUR <int>, MINUTE <int>, NHS <int>, ROAD_FNC <int>,
## #   ROUTE <int>, TWAY_ID <chr>, TWAY_ID2 <chr>, MILEPT <int>,
## #   LATITUDE <dbl>, LONGITUD <dbl>, SP_JUR <int>, HARM_EV <int>,
## #   MAN_COLL <int>, RELJCT1 <int>, RELJCT2 <int>, TYP_INT <int>,
## #   WRK_ZONE <int>, REL_ROAD <int>, LGT_COND <int>, WEATHER1 <int>,
## #   WEATHER2 <int>, WEATHER <int>, SCH_BUS <int>, RAIL <chr>,
## #   NOT_HOUR <int>, NOT_MIN <int>, ARR_HOUR <int>, ARR_MIN <int>,
## #   HOSP_HR <int>, HOSP_MN <int>, CF1 <int>, CF2 <int>, CF3 <int>,
## #   FATALS <int>, DRUNK_DR <int>
```

fars_read function

This function is used to load in the data.

```
fars_read("accident_2013.csv.bz2")
```

```
## # A tibble: 30,202 x 50
##    STATE ST_CASE VE_TOTAL VE_FORMS PVH_INVL  PEDS PERNOTMVIT PERMVIT
##    <int>   <int>    <int>    <int>    <int> <int>      <int>   <int>
##  1     1   10001        1        1        0     0          0       8
##  2     1   10002        2        2        0     0          0       2
##  3     1   10003        1        1        0     0          0       1
##  4     1   10004        1        1        0     0          0       3
##  5     1   10005        2        2        0     0          0       3
##  6     1   10006        2        2        0     0          0       3
```

```
##  7     1   10007        1        1        0     0          0       1
##  8     1   10008        2        2        0     0          0       2
##  9     1   10009        1        1        0     0          0       1
## 10     1   10010        2        2        0     0          0       4
## # ... with 30,192 more rows, and 42 more variables: PERSONS <int>,
## #   COUNTY <int>, CITY <int>, DAY <int>, MONTH <int>, YEAR <int>,
## #   DAY_WEEK <int>, HOUR <int>, MINUTE <int>, NHS <int>, ROAD_FNC <int>,
## #   ROUTE <int>, TWAY_ID <chr>, TWAY_ID2 <chr>, MILEPT <int>,
## #   LATITUDE <dbl>, LONGITUD <dbl>, SP_JUR <int>, HARM_EV <int>,
## #   MAN_COLL <int>, RELJCT1 <int>, RELJCT2 <int>, TYP_INT <int>,
## #   WRK_ZONE <int>, REL_ROAD <int>, LGT_COND <int>, WEATHER1 <int>,
## #   WEATHER2 <int>, WEATHER <int>, SCH_BUS <int>, RAIL <chr>,
## #   NOT_HOUR <int>, NOT_MIN <int>, ARR_HOUR <int>, ARR_MIN <int>,
## #   HOSP_HR <int>, HOSP_MN <int>, CF1 <int>, CF2 <int>, CF3 <int>,
## #   FATALS <int>, DRUNK_DR <int>
```

make_filename

The function is used to create a string as the name of the file. It uses the year as suffix.

```
make_filename(2012)
```

```
## [1] "accident_2012.csv.bz2"
```

fars_read_years

This function is used to grab the month and the year from data.

```
fars_read_years(2013:2014)
```

```
## [[1]]
## # A tibble: 30,202 x 2
##    MONTH  year
##    <int> <int>
##  1     1  2013
##  2     1  2013
##  3     1  2013
##  4     1  2013
##  5     1  2013
##  6     1  2013
##  7     1  2013
##  8     1  2013
##  9     1  2013
## 10     1  2013
## # ... with 30,192 more rows
##
## [[2]]
## # A tibble: 30,056 x 2
##    MONTH  year
##    <int> <int>
##  1     1  2014
##  2     1  2014
##  3     1  2014
##  4     1  2014
```

```
##  5     1  2014
##  6     1  2014
##  7     1  2014
##  8     1  2014
##  9     1  2014
## 10     1  2014
## # ... with 30,046 more rows
```

fars_summarize_years

The function is used to get a data frame containing the total number of observations in each month for each year. It counted all observations in the data.

```
fars_summarize_years(2013:2014)
```

```
## # A tibble: 12 x 3
##    MONTH `2013` `2014`
##    <int>  <int>  <int>
##  1     1   2230   2168
##  2     2   1952   1893
##  3     3   2356   2245
##  4     4   2300   2308
##  5     5   2532   2596
##  6     6   2692   2583
##  7     7   2660   2696
##  8     8   2899   2800
##  9     9   2741   2618
## 10    10   2768   2831
## 11    11   2615   2714
## 12    12   2457   2604
```

fars_map_state

The function is used to draw a map showing the accident rate for each state. The state number is from 1 to 51.

```
unique(dt$STATE)
```

```
##  [1]  1  2  4  5  6  8  9 10 11 12 13 15 16 17 18 19 20 21 22 23 24 25 26
## [24] 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 44 45 46 47 48 49 50
## [47] 51 53 54 55 56
```

```
fars_map_state(1, 2014)
```