Homework 3

Evan Yacek ety78

This homework is due on Feb. 9, 2016 at 11:59pm. Please submit as a PDF file on Canvas.

In this homework, you are asked to evaluate two data sets and determine if they are tidy data sets. *We are referring to a very specific definition of "tidy", so if this term is unfamiliar to you, please review the lecture materials*.

Question 1: (4 pts) *The dataset* `VADeaths` *built into R lists death rates per 1000 people in Virginia in 1940. You can run* `?VADeaths` *to learn more about this data set.*

```
VADeaths
```

```
##       Rural Male Rural Female Urban Male Urban Female
## 50-54       11.7          8.7       15.4          8.4
## 55-59       18.1         11.7       24.3         13.6
## 60-64       26.9         20.3       37.0         19.3
## 65-69       41.0         30.9       54.6         35.1
## 70-74       66.0         54.3       71.1         50.0
```

NO the data is not tidy. Each variable is not saved in its own column, Their should be a column labeled age group.

The data set `InsectSprays` *built into R contains counts of insects (in agricultural units) after treatment with different types of insecticides. You should be familiar with this data set from Homework 1. You can run* `?InsectSprays` *to learn more about this data set.*

```
head(InsectSprays)
```

```
##   count spray
## 1    10     A
## 2     7     A
## 3    20     A
## 4    14     A
## 5    14     A
## 6    12     A
```

Using the formal definition of tidy data that we learned in lecture, is this data set tidy? Explain why or why not.

Yes InsectSprays is tidy. Each variable forms a column, each observation forms a row, and finally each observational unit forms a table.

Question 2: (2 pts) *The* `gapminder` *dataset from the "gapminder" package contains information about life expectancy, GDP per capita, and population by country from 1952 to 2007.* ***NOTE:*** *You will have to install the 'gapminder' package using the command* `install.packages("gapminder")` *before you can load the* `gapminder` *dataset.*

```
library(gapminder)
```

```
## Warning: package 'gapminder' was built under R version 3.1.3
```

```
head(gapminder)
```

```
## Source: local data frame [6 x 6]
##
##       country continent  year lifeExp      pop gdpPercap
##        (fctr)    (fctr) (int)   (dbl)    (int)     (dbl)
## 1 Afghanistan      Asia  1952  28.801  8425333  779.4453
## 2 Afghanistan      Asia  1957  30.332  9240934  820.8530
## 3 Afghanistan      Asia  1962  31.997 10267083  853.1007
## 4 Afghanistan      Asia  1967  34.020 11537966  836.1971
## 5 Afghanistan      Asia  1972  36.088 13079460  739.9811
## 6 Afghanistan      Asia  1977  38.438 14880372  786.1134
```

*Pick a continent and a year of your choosing. What is the **median** life expectancy for people living on that continent in the year that you chose? State your answer in a sentence. **NOTE:** The `gapminder` data set contains life expectancy data in 5 year increments beginning in 1952. Also, do not worry about the fact that you are computing a median of medians and not the true median life expectancy for the whole continent.*

I chose Africa in 1977.

```
years <- gapminder[gapminder$year == 1977,]
Africa77 <- years[years$continent == "Africa",]
Africa77
```

```
## Source: local data frame [52 x 6]
##
##                     country continent  year lifeExp      pop gdpPercap
##                      (fctr)    (fctr) (int)   (dbl)    (int)     (dbl)
## 1                   Algeria    Africa  1977  58.014 17152804 4910.4168
## 2                    Angola    Africa  1977  39.483  6162675 3008.6474
## 3                     Benin    Africa  1977  49.190  3168267 1029.1613
## 4                  Botswana    Africa  1977  59.319   781472 3214.8578
## 5              Burkina Faso    Africa  1977  46.137  5889574  743.3870
## 6                   Burundi    Africa  1977  45.910  3834415  556.1033
## 7                  Cameroon    Africa  1977  49.355  7959865 1783.4329
## 8  Central African Republic    Africa  1977  46.775  2167533 1109.3743
## 9                      Chad    Africa  1977  47.383  4388260 1133.9850
## 10                  Comoros    Africa  1977  50.939   304739 1172.6030
## ..                      ...       ...   ...     ...      ...       ...
```

```
median(Africa77$lifeExp)
```

```
## [1] 49.2725
```

The median life expectancy in Africa in 1977 is 49.27.

Question 3: (4 pts) *Which countries experienced the largest change in life expectancy between **1988 and 2007**? List at least the top 5 and state your answer in a sentence. **HINT:** Use the functions* `max()` *and* `min()` *to determine the net change in life expectancy.*

```
LargestChange <- gapminder %>% filter(year >= 1988 & year <= 2007) %>% group_by(country) %>% summariz
e(LifeE = max(lifeExp)-min(lifeExp)) %>% arrange(desc(LifeE)) %>% slice(1:5)
LargestChange
```

```
## Source: local data frame [5 x 2]
##
##     country  LifeE
##      (fctr)  (dbl)
## 1    Rwanda 22.643
## 2  Zimbabwe 20.388
## 3 Swaziland 18.861
## 4   Lesotho 17.093
## 5  Botswana 16.111
```

Rwanda and Zimbabwe had the largest net change in life expectancy between 1988 and 2007. However, Swaiziland, Lesotho, and Botswana followed close behind.

*Choose a country from your list above and create a line plot of the life expectancy over time between **1952 and 2007**.*

I chose Zimbabwe

```
gapminderfiltered <- filter( gapminder, year >= 1952 & year <= 2007)
zimData <- gapminderfiltered[gapminderfiltered$country == "Zimbabwe",]
zimData
```

```
## Source: local data frame [12 x 6]
##
##     country continent  year lifeExp      pop gdpPercap
##      (fctr)    (fctr) (int)   (dbl)    (int)     (dbl)
## 1  Zimbabwe    Africa  1952  48.451  3080907  406.8841
## 2  Zimbabwe    Africa  1957  50.469  3646340  518.7643
## 3  Zimbabwe    Africa  1962  52.358  4277736  527.2722
## 4  Zimbabwe    Africa  1967  53.995  4995432  569.7951
## 5  Zimbabwe    Africa  1972  55.635  5861135  799.3622
## 6  Zimbabwe    Africa  1977  57.674  6642107  685.5877
## 7  Zimbabwe    Africa  1982  60.363  7636524  788.8550
## 8  Zimbabwe    Africa  1987  62.351  9216418  706.1573
## 9  Zimbabwe    Africa  1992  60.377 10704340  693.4208
## 10 Zimbabwe    Africa  1997  46.809 11404948  792.4500
## 11 Zimbabwe    Africa  2002  39.989 11926563  672.0386
## 12 Zimbabwe    Africa  2007  43.487 12311143  469.7093
```

```
ggplot(data = zimData, aes(x = year, y = lifeExp))+geom_line()
```