

MA615/415 Midterm Project

County-level oil and gas production

liuchenlu

Set Up: Packages and Functions

We will need to use the following packages: "stringr", "ggplot2", and "dplyr". install them first (outside this Rmd).

```
install.packages("stringr")
```

```
install.packages("ggplot2")
```

```
install.packages("dplyr")
```

```
# use this code chunk to load all the packages that you will be using  
library(dplyr)
```

```
##  
## 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(plyr)
```

```
## -----  
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)  
## -----  
##  
## Attaching package: 'plyr'  
##  
## The following objects are masked from 'package:dplyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize
```

```
library(ggplot2)  
# (do not include commands to install the packages)
```

Import Raw Data in R

I already downloaded a copy of the CSV files with the raw data :

- oilgascounty.csv

Now, use one of the reading table functions to import the data sets in R. Do NOT convert strings into factors.

```
# read in oilgascounty.csv
oilgascounty = read.csv("D:\\midterm\\oilgascounty.csv", stringsAsFactors = FALSE)

# check structure with str()

str(oilgascounty)
```

```
## 'data.frame':    3109 obs. of  35 variables:
##  $ FIPS                : int  1001 1003 1005 1007 1009 1011 1013 1015 1017 1019 ...
##  $ geoid                : int  1001 1003 1005 1007 1009 1011 1013 1015 1017 1019 ...
##  $ Stabr                : chr  "AL" "AL" "AL" "AL" ...
##  $ County_Name          : chr  "Autauga County" "Baldwin County" "Barbour County" "Bibb Co
##  $ Rural_Urban_Continuum_Code_2013: int  2 3 6 1 1 6 6 3 6 6 ...
##  $ Urban_Influence_2013  : int  2 2 6 1 1 6 6 2 5 6 ...
##  $ Metro_Nonmetro_2013   : int  1 1 0 1 1 0 0 1 0 0 ...
##  $ Metro_Micro_Noncore_2013 : int  2 2 0 2 2 0 0 2 1 0 ...
##  $ oil2000              : int  0 138072 0 0 0 0 0 0 0 0 ...
##  $ oil2001              : int  0 134666 0 0 0 0 0 0 0 0 ...
##  $ oil2002              : int  0 138011 0 0 0 0 0 0 0 0 ...
##  $ oil2003              : int  0 127985 0 0 0 0 0 0 0 0 ...
##  $ oil2004              : int  0 130763 0 0 0 0 0 0 0 0 ...
##  $ oil2005              : int  0 118043 0 0 0 0 0 0 0 0 ...
##  $ oil2006              : int  0 103992 0 0 0 0 0 0 0 0 ...
##  $ oil2007              : int  0 112303 0 0 0 0 0 0 0 0 ...
##  $ oil2008              : int  0 97623 0 0 0 0 0 0 0 0 ...
##  $ oil2009              : int  0 84982 0 0 0 0 0 0 0 0 ...
##  $ oil2010              : int  0 101955 0 0 0 0 0 0 0 0 ...
##  $ oil2011              : int  0 94638 0 0 0 0 0 0 0 0 ...
##  $ gas2000              : int  0 72543902 0 0 0 0 0 0 0 0 ...
##  $ gas2001              : int  0 98699994 0 0 0 0 0 0 0 0 ...
##  $ gas2002              : int  0 107142655 0 0 0 0 0 0 0 0 ...
##  $ gas2003              : int  0 101510068 0 0 0 0 0 0 0 0 ...
##  $ gas2004              : int  0 90146850 0 0 0 0 0 0 0 0 ...
##  $ gas2005              : int  0 84536875 0 8301 0 0 0 0 0 0 ...
##  $ gas2006              : int  0 83951640 0 98853 0 0 0 0 0 0 ...
##  $ gas2007              : int  0 82876786 0 480015 0 0 0 0 0 0 ...
##  $ gas2008              : int  0 78547145 0 684143 20516 0 0 0 0 0 ...
##  $ gas2009              : int  0 68525628 0 551719 61054 0 0 0 0 0 ...
##  $ gas2010              : int  0 63069025 0 453132 3594 0 0 0 0 0 ...
##  $ gas2011              : int  0 51041072 0 400504 21496 0 0 0 0 0 ...
##  $ oil_change_group      : chr  "Status Quo" "Status Quo" "Status Quo" "Status Quo" ...
##  $ gas_change_group      : chr  "Status Quo" "H_Decline" "Status Quo" "Status Quo" ...
##  $ oil_gas_change_group  : chr  "Status Quo" "H_Decline" "Status Quo" "Status Quo" ...
```

FIPS: Five digit Federal Information Processing Standard (FIPS) code (numeric)

geoid: FIPS code with leading zero (string)

Stabr: State abbreviation (string)

County_Name: County name (string)

Rural_Urban_Continuum_Code_2013: Rural-urban Continuum Code 2013 (see code descriptions)

Urban_Influence_2013: Urban Influence Code, 2013 (see code descriptions)

Metro_Nonmetro_2013: Metro-nonmetro 2013 (0=nonmetro, 1=metro)

Metro_Micro_Noncore_2013: Metro-Micro-Noncore Indicator 2013 (0=nonmetro noncore, 1=nonmetro micropolitan, 2=metropolitan)

oil2000, oil2001, ..., oil2011: Annual gross withdrawals (barrels) of crude oil, for the year specified in the variable name

gas2000, gas2001, ..., gas2011: Annual gross withdrawals (thousand cubic feet) of natural gas, for the year specified in the variable name

oil_change_group: Categorical variable based upon change in the dollar value of oil production, 2000-11. Values are H_Growth (\geq \$20 million), H_Decline (\leq -\$20 million), Status Quo (change between +/- \$20 million)

gas_change_group: Categorical variable based upon change in the dollar value of natural gas production, 2000-11. Values are H_Growth (\geq \$20 million), H_Decline (\leq -\$20 million), Status Quo (change between +/- \$20 million)

oil_gas_change_group: Categorical variable based on the change in the dollar value of the sum of oil and natural gas production, 2000-11. Values are H_Growth (\geq \$20 million), H_Decline (\leq -\$20 million), Status Quo (change between +/- \$20 million)

Cleaning raw data of County-level oil and gas production

First, we summary the data

```
summary(oilgascounty)
```

```
##      FIPS      geoid      Stabr      County_Name
## Min.   : 1001  Min.   : 1001  Length:3109  Length:3109
## 1st Qu.:19045  1st Qu.:19045  Class :character  Class :character
## Median :29213  Median :29213  Mode  :character  Mode  :character
## Mean   :30679  Mean   :30679
## 3rd Qu.:46009  3rd Qu.:46009
## Max.   :56045  Max.   :56045
## Rural_Urban_Continuum_Code_2013 Urban_Influence_2013 Metro_Nonmetro_2013
## Min.   :1.000      Min.   : 1.000      Min.   :0.0000
## 1st Qu.:2.000      1st Qu.: 2.000      1st Qu.:0.0000
## Median :6.000      Median : 5.000      Median :0.0000
## Mean   :4.986      Mean   : 5.224      Mean   :0.3734
## 3rd Qu.:7.000      3rd Qu.: 8.000      3rd Qu.:1.0000
## Max.   :9.000      Max.   :12.000      Max.   :1.0000
## Metro_Micro_Noncore_2013 oil2000      oil2001
## Min.   :0.0000      Min.   :      0      Min.   :      0
## 1st Qu.:0.0000      1st Qu.:      0      1st Qu.:      0
## Median :1.0000      Median :      0      Median :      0
```

| | | |
|----------------------|-------------------|-------------------|
| ## Mean :0.9518 | Mean : 389916 | Mean : 377786 |
| ## 3rd Qu.:2.0000 | 3rd Qu.: 7740 | 3rd Qu.: 8115 |
| ## Max. :2.0000 | Max. :208781424 | Max. :200867504 |
| ## oil2002 | oil2003 | oil2004 |
| ## Min. : 0 | Min. : 0 | Min. : 0 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.: 0 |
| ## Median : 0 | Median : 0 | Median : 0 |
| ## Mean : 364100 | Mean : 356800 | Mean : 350380 |
| ## 3rd Qu.: 8520 | 3rd Qu.: 7802 | 3rd Qu.: 8566 |
| ## Max. :198730576 | Max. :191390816 | Max. :185325184 |
| ## oil2005 | oil2006 | oil2007 |
| ## Min. : 0 | Min. : 0 | Min. : 0 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.: 0 |
| ## Median : 0 | Median : 0 | Median : 0 |
| ## Mean : 346858 | Mean : 348923 | Mean : 351030 |
| ## 3rd Qu.: 9468 | 3rd Qu.: 9785 | 3rd Qu.: 10633 |
| ## Max. :177573184 | Max. :170139232 | Max. :166179936 |
| ## oil2008 | oil2009 | oil2010 |
| ## Min. : 0 | Min. : 0 | Min. : 0 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.: 0 |
| ## Median : 0 | Median : 0 | Median : 0 |
| ## Mean : 362588 | Mean : 360793 | Mean : 379168 |
| ## 3rd Qu.: 12365 | 3rd Qu.: 11014 | 3rd Qu.: 12181 |
| ## Max. :162249248 | Max. :154649312 | Max. :147378048 |
| ## oil2011 | gas2000 | gas2001 |
| ## Min. : 0 | Min. : 0 | Min. : 0 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.: 0 |
| ## Median : 0 | Median : 0 | Median : 0 |
| ## Mean : 432846 | Mean : 5057060 | Mean : 5143353 |
| ## 3rd Qu.: 12427 | 3rd Qu.: 34653 | 3rd Qu.: 34457 |
| ## Max. :140593024 | Max. :706343704 | Max. :673492347 |
| ## gas2002 | gas2003 | gas2004 |
| ## Min. : 0 | Min. : 0 | Min. : 0 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.: 0 |
| ## Median : 0 | Median : 0 | Median : 0 |
| ## Mean : 5105789 | Mean : 5154097 | Mean : 5306609 |
| ## 3rd Qu.: 32729 | 3rd Qu.: 30875 | 3rd Qu.: 41730 |
| ## Max. :637562208 | Max. :655573171 | Max. :731254191 |
| ## gas2005 | gas2006 | gas2007 |
| ## Min. : 0 | Min. : 0 | Min. :0.000e+00 |
| ## 1st Qu.: 0 | 1st Qu.: 0 | 1st Qu.:0.000e+00 |
| ## Median : 0 | Median : 0 | Median :0.000e+00 |
| ## Mean : 5338176 | Mean : 5590905 | Mean :5.862e+06 |
| ## 3rd Qu.: 42899 | 3rd Qu.: 53689 | 3rd Qu.:6.239e+04 |
| ## Max. :815025697 | Max. :880441069 | Max. :1.010e+09 |
| ## gas2008 | gas2009 | gas2010 |
| ## Min. :0.000e+00 | Min. :0.000e+00 | Min. :0.000e+00 |
| ## 1st Qu.:0.000e+00 | 1st Qu.:0.000e+00 | 1st Qu.:0.000e+00 |
| ## Median :0.000e+00 | Median :0.000e+00 | Median :0.000e+00 |
| ## Mean :6.398e+06 | Mean :6.527e+06 | Mean :6.794e+06 |
| ## 3rd Qu.:7.222e+04 | 3rd Qu.:6.710e+04 | 3rd Qu.:6.216e+04 |
| ## Max. :1.145e+09 | Max. :1.193e+09 | Max. :1.198e+09 |
| ## gas2011 | oil_change_group | gas_change_group |
| ## Min. :0.000e+00 | Length:3109 | Length:3109 |

```
## 1st Qu.:0.000e+00 Class :character Class :character
## Median :0.000e+00 Mode :character Mode :character
## Mean :7.430e+06
## 3rd Qu.:6.045e+04
## Max. :1.167e+09
## oil_gas_change_group
## Length:3109
## Class :character
## Mode :character
##
##
##
```

then we count the number of `oil_gas_change_group` in each (`Metro_Nonmetro_2013`) with the tidy `count` function.

```
count(oilgascounty, c("Metro_Nonmetro_2013", "oil_gas_change_group"))
```

```
## Metro_Nonmetro_2013 oil_gas_change_group freq
## 1 0 H_Decline 143
## 2 0 H_Growth 163
## 3 0 Status Quo 1642
## 4 1 H_Decline 69
## 5 1 H_Growth 55
## 6 1 Status Quo 1037
```

count the number of `oil_gas_change_group` in each (`Metro_Micro_Noncore_2013`) with the tidy `count` function.

```
count(oilgascounty, c("Metro_Micro_Noncore_2013", "oil_gas_change_group"))
```

```
## Metro_Micro_Noncore_2013 oil_gas_change_group freq
## 1 0 H_Decline 99
## 2 0 H_Growth 114
## 3 0 Status Quo 1098
## 4 1 H_Decline 44
## 5 1 H_Growth 49
## 6 1 Status Quo 544
## 7 2 H_Decline 69
## 8 2 H_Growth 55
## 9 2 Status Quo 1037
```

Then we remove missing (and hence uninformative for our purpose) values with `subset`.

```
oil_group2 <- subset(oilgascounty, oil2000>0&oil2001>0&oil2002>0&oil2003>0&oil2004>0&oil2005>0
&oil2006>0&oil2007>0&oil2008>0&oil2009>0&oil2010>0&oil2011>0)

oil_group3=apply(oil_group2[,9:20],2,sum)
new_data=data.frame(time=2000:2011,stype=rep("oil",12),prudction=as.numeric(oil_group3))
gas_group2 <- subset(oilgascounty, gas2000>0&gas2001>0&gas2002>0&gas2003>0&gas2004>0&gas2005>0
&gas2006>0&gas2007>0&gas2008>0&gas2009>0&gas2010>0&gas2011>0)
gas_group3=apply(gas_group2[,9:20],2,sum)
new_data=rbind(new_data,data.frame(time=2000:2011,stype=rep("gas",12),prudction=as.numeric(gas_group3)))
```

Data Visualization

Use the `records` data frame, and functions in "ggplot2", to create charts similar to those displayed in the PDF with the instructions for this project:

```
qplot(time, prudction, data = new_data, geom=c("line", "point"),  
      main="oil and gas Tidy Data ", colour = stype)
```

