Problem Set

1 Preparations

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.3
```

```
## -- Attaching packages -------- tidyverse 1.2.1 --
```

```
## v ggplot2 2.2.1     v purrr   0.2.4
## v tibble  1.4.1     v dplyr   0.7.4
## v tidyr   0.7.2     v stringr 1.2.0
## v readr   1.1.1     v forcats 0.2.0
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
## Warning: package 'tibble' was built under R version 3.4.3
```

```
## Warning: package 'tidyr' was built under R version 3.4.3
```

```
## Warning: package 'readr' was built under R version 3.4.3
```

```
## Warning: package 'purrr' was built under R version 3.4.3
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
## Warning: package 'stringr' was built under R version 3.4.3
```

```
## Warning: package 'forcats' was built under R version 3.4.3
```

```
## -- Conflicts ----------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

1.1 Data Import

In this problem, I tried working with Penn World Table ver. 9.0.

```
pwt <- haven::read_dta("pwt90.dta")
pwt
```

```
## # A tibble: 11,830 x 47
##    coun~ coun~ curr~  year rgdpe rgdpo   pop   emp   avh    hc  ccon   cda
##    <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 ABW   Aruba Arub~  1950    NA    NA    NA    NA    NA    NA    NA    NA
##  2 ABW   Aruba Arub~  1951    NA    NA    NA    NA    NA    NA    NA    NA
##  3 ABW   Aruba Arub~  1952    NA    NA    NA    NA    NA    NA    NA    NA
##  4 ABW   Aruba Arub~  1953    NA    NA    NA    NA    NA    NA    NA    NA
##  5 ABW   Aruba Arub~  1954    NA    NA    NA    NA    NA    NA    NA    NA
##  6 ABW   Aruba Arub~  1955    NA    NA    NA    NA    NA    NA    NA    NA
##  7 ABW   Aruba Arub~  1956    NA    NA    NA    NA    NA    NA    NA    NA
##  8 ABW   Aruba Arub~  1957    NA    NA    NA    NA    NA    NA    NA    NA
##  9 ABW   Aruba Arub~  1958    NA    NA    NA    NA    NA    NA    NA    NA
## 10 ABW   Aruba Arub~  1959    NA    NA    NA    NA    NA    NA    NA    NA
```

```
## # ... with 11,820 more rows, and 35 more variables: cgdpe <dbl>, cgdpo
## #   <dbl>, ck <dbl>, ctfp <dbl>, cwtfp <dbl>, rgdpna <dbl>, rconna <dbl>,
## #   rdana <dbl>, rkna <dbl>, rtfpna <dbl>, rwtfpna <dbl>, labsh <dbl>,
## #   delta <dbl>, xr <dbl>, pl_con <dbl>, pl_da <dbl>, pl_gdpo <dbl>, i_cig
## #   <dbl+lbl>, i_xm <dbl+lbl>, i_xr <dbl+lbl>, i_outlier <dbl+lbl>,
## #   cor_exp <dbl>, statcap <dbl>, csh_c <dbl>, csh_i <dbl>, csh_g <dbl>,
## #   csh_x <dbl>, csh_m <dbl>, csh_r <dbl>, pl_c <dbl>, pl_i <dbl>, pl_g
## #   <dbl>, pl_x <dbl>, pl_m <dbl>, pl_k <dbl>
```

1.2 Data Visualization

```
countries <- c("United States", "united Kingdom", "Germany", "France", "Italy", "Japan", "Canada", "Chi
pwt10 <- pwt%>%
  filter(country %in% countries) %>%
  select(country, year, rgdpo, pop)
```

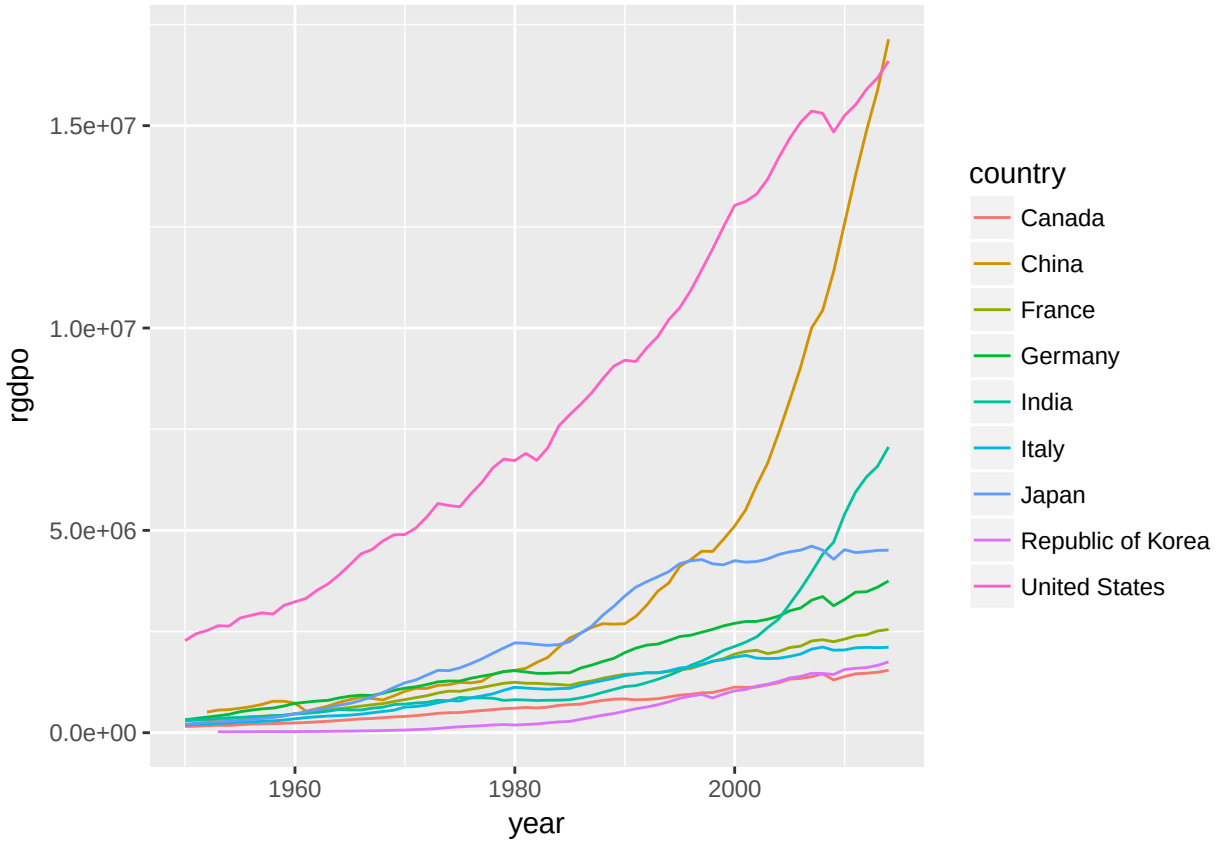
```
## Warning: package 'bindrcpp' was built under R version 3.4.3
```

```
pwt10
```

```
## # A tibble: 585 x 4
##    country  year  rgdpo   pop
##    <chr>   <dbl>  <dbl> <dbl>
##  1 Canada   1950 155053  13.8
##  2 Canada   1951 160307  14.1
##  3 Canada   1952 174148  14.6
##  4 Canada   1953 182327  15.0
##  5 Canada   1954 181437  15.4
##  6 Canada   1955 197522  15.8
##  7 Canada   1956 213976  16.2
##  8 Canada   1957 219339  16.8
##  9 Canada   1958 224430  17.2
## 10 Canada   1959 233374  17.6
## # ... with 575 more rows
```

```
ggplot(pwt10) + geom_line(aes(x=year, y = rgdpo, color = country))
```

```
## Warning: Removed 5 rows containing missing values (geom_path).
```



1.3 Data Transformation

```
tbl <- tibble(
  id = letters[1:4],
  salary = 400 + rnorm(4,0,50),
  sex = c("M", "M", "F", "F")
)
tbl
```

```
## # A tibble: 4 x 3
##   id    salary sex
##   <chr>  <dbl> <chr>
## 1 a        326 M
## 2 b        359 M
## 3 c        481 F
## 4 d        455 F
```

```
mutate(tbl, salary_in_thousand = 1000*salary)
```

```
## # A tibble: 4 x 4
##   id    salary sex   salary_in_thousand
##   <chr>  <dbl> <chr>               <dbl>
## 1 a        326 M                  325598
## 2 b        359 M                  358984
## 3 c        481 F                  481129
```

```
## 4 d        455 F                 454987
```