Lab_12

Fantastic Four

Functions Defined

```
#for defining functions (?):
showLevels <- function(data_input){
  return(unique(data_input))
}

summaryStats <- function(data_input,continent_in,start_year,end_year,column_in){
  #main function to return the summary stats for a certain dataset and column for years
  x2 <- filter(data_input,continent == continent_in,year>=start_year,year<=end_year)
  x <- sample_n(x2,10,replace=TRUE)
  df <- data.frame(continent_in,start_year,end_year,mean(x[[column_in]]),median(x[[column_in]]),min(x[[
  return(df)
}

calc_summaryStats <- function(data_input,continent_in,start_year,end_year,column_in){
  yearA <- start_year
  output_summary <- 0

  while(yearA < end_year){
    output_summary <- rbind(output_summary, summaryStats(data_input,continent_in,yearA,yearA+9,column_i
    yearA <- yearA + 10
  }
  output_summary <- output_summary[-1,]
  colnames(output_summary) <- c("continent","start_year","end_year","mean","median","minimum","maximum"
  return(output_summary)
}

runAll_calc_summaryStats <- function(data_input,start_year,end_year,column_in){
  #cont = c("Africa","Asia","Americas","Europe","Oceania")
  output_dataframe <- 0
  fun <- showLevels(data_input$continent)
  for(item in fun){
    output_dataframe <- rbind(output_dataframe, calc_summaryStats(gapminder,item,start_year,end_year,co
  }
  return(output_dataframe)

}
```

Main part

```
# Is there a correlation between life expectancy and population size?
runAll_calc_summaryStats(gapminder,1950,2010,"pop")
```

```
## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
```

```
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated

## Warning in `[<-.factor`(`*tmp*`, ri, value = 0): invalid factor level, NA
## generated
```

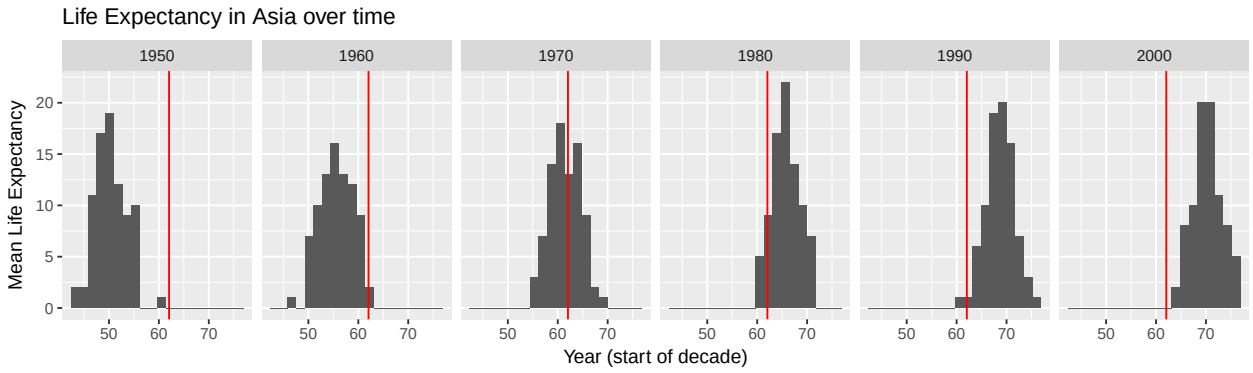
	continent	start_year	end_year	mean	median	minimum	maximum
1	<NA>	0	0	0	0.0	0	0
2	Asia	1950	1959	80855051	29309280.0	1944401	556263527
3	Asia	1960	1969	150194190	6730039.0	171863	774510000
4	Asia	1970	1979	121289142	15946388.0	829050	916395000
5	Asia	1980	1989	118336805	8465051.5	1593882	981235000
6	Asia	1990	1999	111300819	8323006.0	273701	959000000
7	Asia	2000	2009	23698974	20868672.0	2713462	62806748
21	Europe	1950	1959	8653568	4354810.0	165110	50127000
31	Europe	1960	1969	11605061	4513084.5	316845	46189000
41	Europe	1970	1979	8054483	9368050.0	4072517	10364869
51	Europe	1980	1989	16681191	7567194.5	364400	56733833
61	Europe	1990	1999	21499300	13841384.0	421014	57374179
71	Europe	2000	2009	38951124	39421694.5	281043	82400996
22	Africa	1950	1959	5300834	4355931.5	170928	16151549
32	Africa	1960	1969	8090155	3339925.0	127617	47287752
42	Africa	1970	1979	7522269	5368696.5	86796	30770372
52	Africa	1980	1989	18253612	7459594.5	304586	81551520
62	Africa	1990	1999	10878470	5048928.5	145608	32160729
72	Africa	2000	2009	16009016	6284238.0	170372	55379852
23	Americas	1950	1959	39212870	4464834.0	1165790	177830000
33	Americas	1960	1969	34489782	19173015.0	97594	180671000
43	Americas	1970	1979	30507726	10158658.5	163778	222585000
53	Americas	1980	1989	5037191	2685879.5	176347	27764644
63	Americas	1990	1999	66963824	28117600.0	264085	263435673
73	Americas	2000	2009	17211599	8884737.0	295131	40301927
24	FSU	1950	1959	75872031	101936816.0	36774854	101936816
34	FSU	1960	1969	2972674	2300376.0	2253604	4564785
44	FSU	1970	1979	29718842	4016013.0	2401244	132556176
54	FSU	1980	1989	2610728	2626695.0	2536976	2664068
64	FSU	1990	1999	29512776	10360516.0	2393657	148446239
74	FSU	2000	2009	45015261	19073389.0	3575439	145266326
25	Oceania	1950	1959	5642531	5460309.5	1908310	10131729
35	Oceania	1960	1969	7382897	10480043.5	421869	12379384

```
## 45   Oceania       1970     1979   2212923    571669.5   106000  13771400
## 55   Oceania       1980     1989   5641227   3154873.0   142207  15788300
## 65   Oceania       1990     1999   8125164   3481477.0    92094  18968247
## 75   Oceania       2000     2009   6818433   2403826.5   196178  19731984
##           IQR
## 1         0.0
## 2  37480497.0
## 3  23171027.2
## 4  85308774.8
## 5  43768839.2
## 6  36662998.0
## 7  15218006.5
## 21  8295339.2
## 31  7108810.8
## 41  4167254.0
## 51  3435605.0
## 61 34575308.5
## 71 50991032.5
## 22  7270396.8
## 32  5371879.0
## 42  5995118.0
## 52 24132650.0
## 62 14018744.2
## 72 28459654.8
## 23 12865719.5
## 33 16663893.5
## 43  9390456.8
## 53  3055797.2
## 63 21032941.8
## 73 26604349.5
## 24 65161962.0
## 34  1730335.5
## 44  5684517.8
## 54    55401.5
## 64 37035397.5
## 74 38047911.0
## 25  6683268.8
## 35  8741858.8
## 45  2206472.8
## 55 11990286.2
## 65 17641490.8
## 75 15060603.0
```

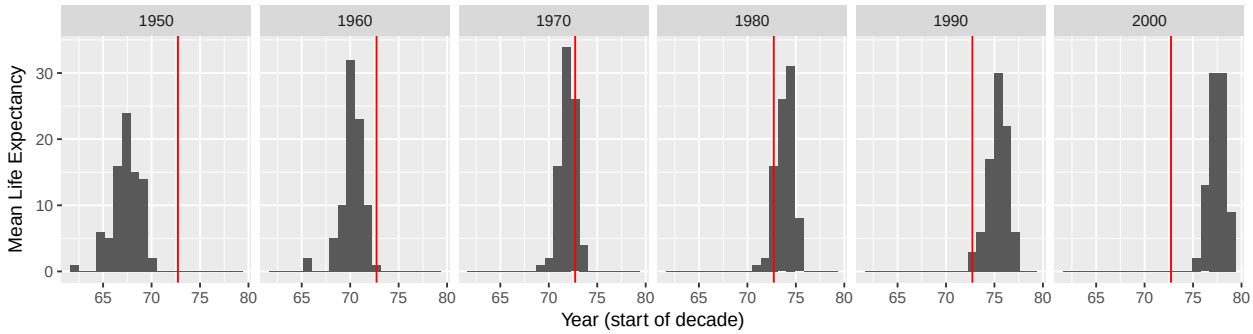
SubQ1

```
# How does life expectancy vary by continent and by decade? Write functions and employ iteration to cal
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Asia",1950,2010,"lifeExp"))
  temp <- temp[-1,]
  i <- i + 1
```

```
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```

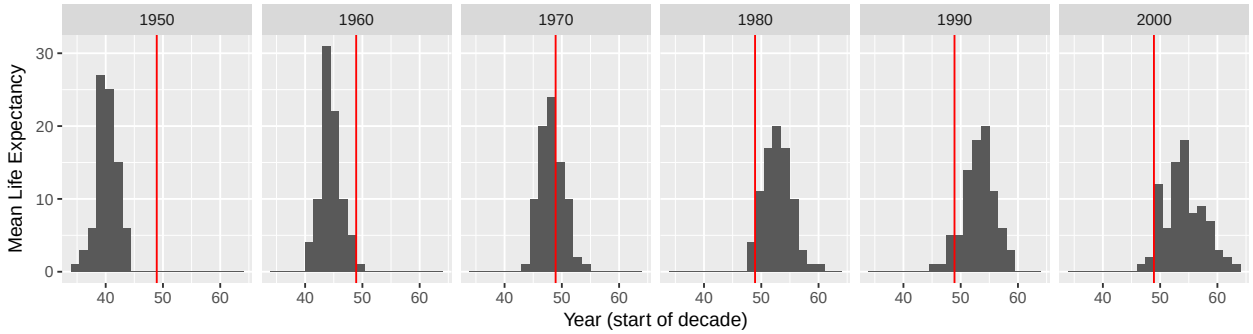



```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Europe",1950,2010,"lifeExp"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```



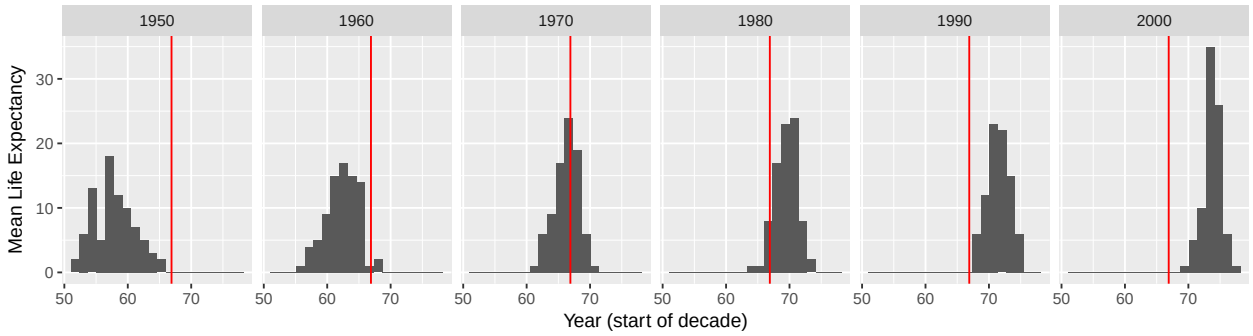

```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Africa",1950,2010,"lifeExp"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```

Life Expectancy in Africa over time

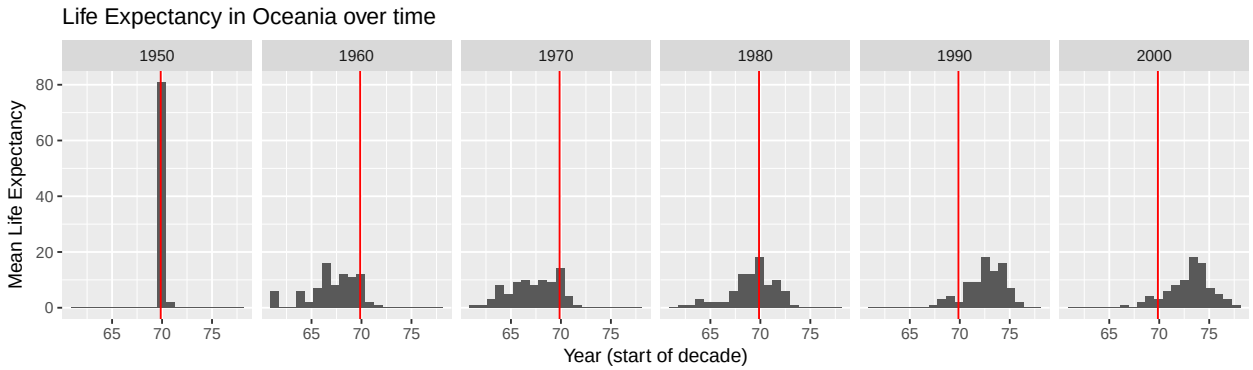



```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Americas",1950,2010,"lifeExp"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```

Life Expectancy in Americas over time



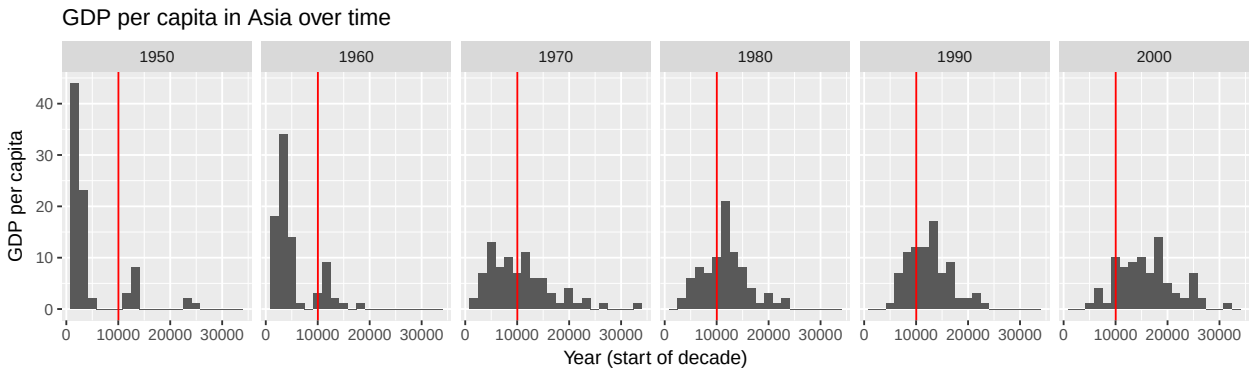

```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Oceania",1950,2010,"lifeExp"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```



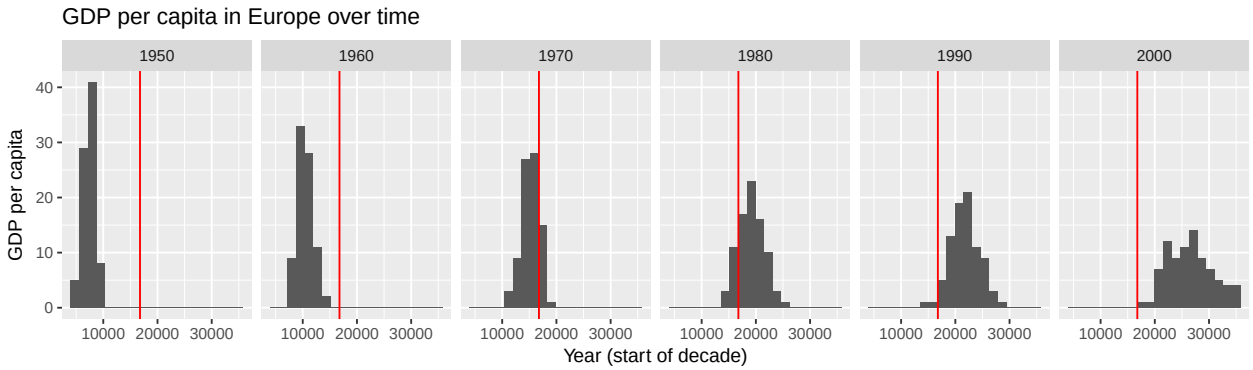
SubQ2

```
# How does GDP per capita vary by continent and by decade? Write functions and employ iteration to calc
```

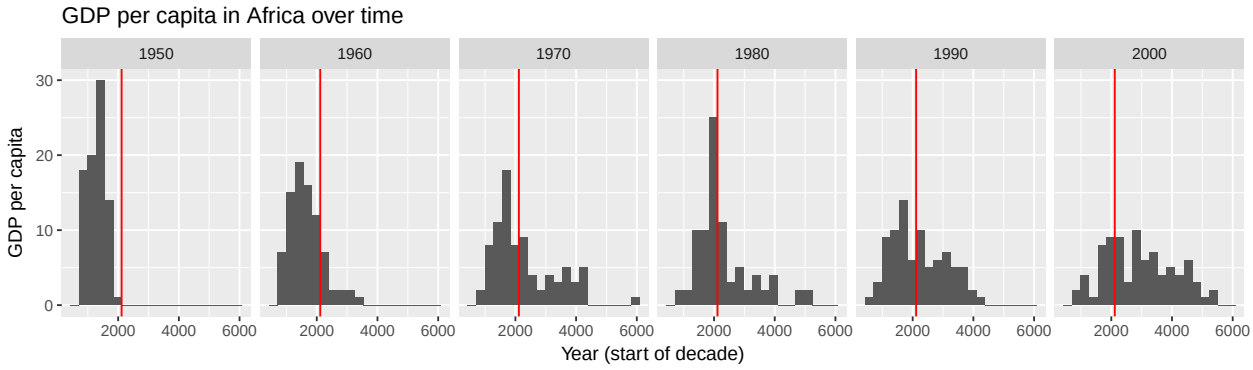
```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Asia",1950,2010,"gdpPercap"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```



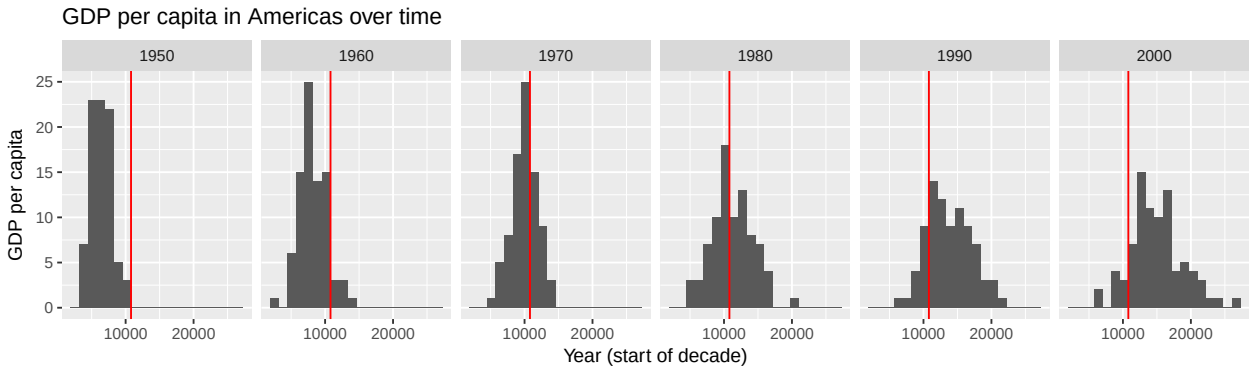

```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Europe",1950,2010,"gdpPercap"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```



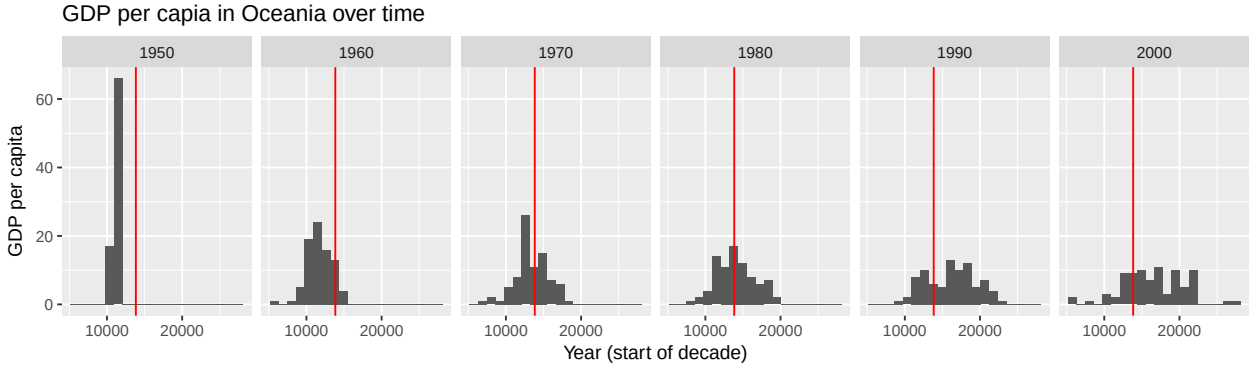

```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Africa",1950,2010,"gdpPercap"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```




```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Americas",1950,2010,"gdpPercap"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```




```
i <- 0
temp <- 0
while(i < 100){
  temp <- rbind(temp,calc_summaryStats(gapminder,"Oceania",1950,2010,"gdpPercap"))
  temp <- temp[-1,]
  i <- i + 1
}

ggplot(data=temp) + geom_histogram(mapping=aes(mean), bins=20) +facet_grid(~start_year) + geom_vline(aes
```



Contributions

- Lindsay:
- Lexie:
- Li:
- Scott: