International Climate Policy - Homework 2 Working document (R Markdown)

Amartya Kumar Sinha

2024-02-01

```
setwd("C:/Users/Amartya Kumar Sinha/OneDrive - The University of Chicago/IntlClimatePolicy_PPHA39930/Assignments/Assn2_IntlClimatePolicy")
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## 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(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library (lfe)
```

```
## Warning: package 'lfe' was built under R version 4.2.3
```

```
## Loading required package: Matrix
```

```
rep_nomiss <- read.csv("icp_indiv_2_dg2011_rep_nomiss.csv")
ave_temp <- read.csv("icp_indiv_2_county_avetemp.csv")
```

Question 1

Taking population weighted average across all temperature variables for the entire country

```r
# Creating a new dataframe where each row is a temperature range and its population-weighted average
temp_vars <- grep("tday", names(rep_nomiss), value = TRUE)

pop_weighted_avg <- sapply(temp_vars, function(x) weighted.mean(rep_nomiss[[x]], rep_nomiss$population, na.rm = TRUE))
```

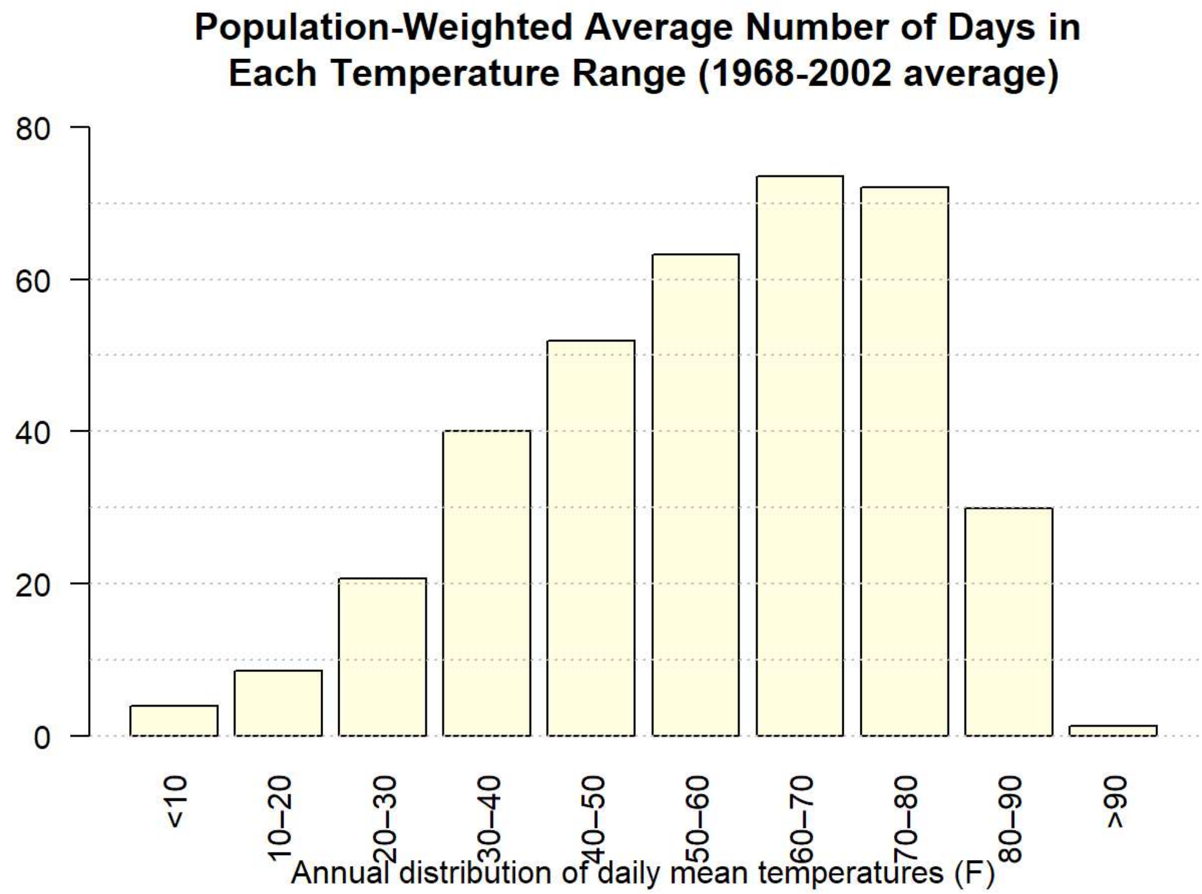
1. a)

Plotting histogram as required

```r
labels <- c("<10", "10-20", "20-30", "30-40", "40-50", "50-60", "60-70", "70-80", "80-90", ">90")

barplot(pop_weighted_avg,
        main = "Population-Weighted Average Number of Days in \nEach Temperature Range (1968-2002 average)",
        ylab = "",
        xlab = "Annual distribution of daily mean temperatures (F)",
        ylim = c(0, 80),
        names.arg = labels,
        las = 2, col = "lightyellow")

abline(h = seq(0, max(pop_weighted_avg), by = 10), col = "gray", lty = "dotted")
```



1.

b) and c) ### Answering questions as required.

```
# Calculating population-weighted average number of days above 90◦ F per year across the US
pop_weighted_avg_above_90 <- weighted.mean(rep_nomiss$tday_gt90, rep_nomiss$population, na.rm = TRUE)

# Finding county with the highest number of days above 90◦ F per year
highest_days_above_90 <- rep_nomiss[which.max(rep_nomiss$tday_gt90), "county"]
highest_days_above_90_count <- max(rep_nomiss$tday_gt90, na.rm = TRUE)

# Calculating the average number of days above 90◦F per year for each county over the sample period (1968-2002)
avg_days_above_90 <- aggregate(tday_gt90 ~ county, rep_nomiss, mean, na.rm = TRUE)

# Finding number of counties that have, on average over the sample period (1968-2002), experienced zero days above 90◦F per year
num_counties_zero_days_above_90 <- sum(avg_days_above_90$tday_gt90 == 0)

# Calculating the total number of counties
total_counties <- length(unique(rep_nomiss$county))

# Calculating the percentage of counties that have, on average over the sample period (1968-2002), experienced zero days above 90◦F per year
percentage_zero_days_above_90 <- (num_counties_zero_days_above_90 / total_counties) * 100

print(paste("The population-weighted average number of days above 90◦ F per year across the US is", round(pop_weighted_avg_above_90, 2)))
```

```
## [1] "The population-weighted average number of days above 90◦ F per year across the US is 1.21"
```

```
print(paste("The county with the highest number of days above 90◦ F per year is", highest_days_above_90, "with", round(highest_days_above_90_count, 2), "days"))
```

```
## [1] "The county with the highest number of days above 90◦ F per year is Yuma County, AZ with 99.54 days"
```

```
print(paste("The number of counties that have, on average over the sample period (1968-2002), experienced zero days above 90◦F per year is", num_counties_zero_days_above_90, "out of a total of", total_counties, "counties in the dataset, or about", round(percentage_zero_days_above_90, 2), "%"))
```

```
## [1] "The number of counties that have, on average over the sample period (1968-2002), experienced zero days above 90◦F per year is 115 out of a total of 2988 counties in the dataset, or about 3.85 %"
```

Question 2

2. a)

```
print(paste("The national average over-65 mortality rate is", round(mean(rep_nomiss$cruderate, n
a.rm = TRUE), 2), "deaths per 100,000 population"))
```

```
## [1] "The national average over-65 mortality rate is 5365.07 deaths per 100,000 population"
```

```
print(paste("The total number of deaths from 1968-2002 is", sum(rep_nomiss$deaths, na.rm = TRU
E)))
```

```
## [1] "The total number of deaths from 1968-2002 is 50830306"
```

Question 3

Merging given datasets and creating variables as instructed

```
# Merging dataframes on 'countycode'
merged_df <- merge(rep_nomiss, ave_temp, by = "countycode")

# Calculating sum of all days over 70◦ F and 80◦ F for each county
merged_df$hotdays <- rowSums(merged_df[,grep("tday_70_80|tday_80_90|tday_gt90", names(merged_d
f))], na.rm = TRUE)
merged_df$hotterdays <- rowSums(merged_df[,grep("tday_80_90|tday_gt90", names(merged_df))], na.r
m = TRUE)

# Calculating  average mortality rate, hotdays, and hotterdays for each county
avg_by_county <- aggregate(cbind(cruderate, hotdays, hotterdays) ~ countycode, merged_df, mean,
na.rm = TRUE)

# Checking first few rows of the new dataframe
print(head(avg_by_county))
```

```
##   countycode cruderate  hotdays hotterdays
## 1       1001  5452.572 151.2064   53.08657
## 2       1003  4698.459 176.5169   72.29507
## 3       1005  5563.602 158.5304   52.39725
## 4       1007  5629.771 144.1877   46.78445
## 5       1009  5106.208 130.5746   32.41731
## 6       1011  5750.395 153.5331   48.80395
```

3. a)

Plotting figure as required

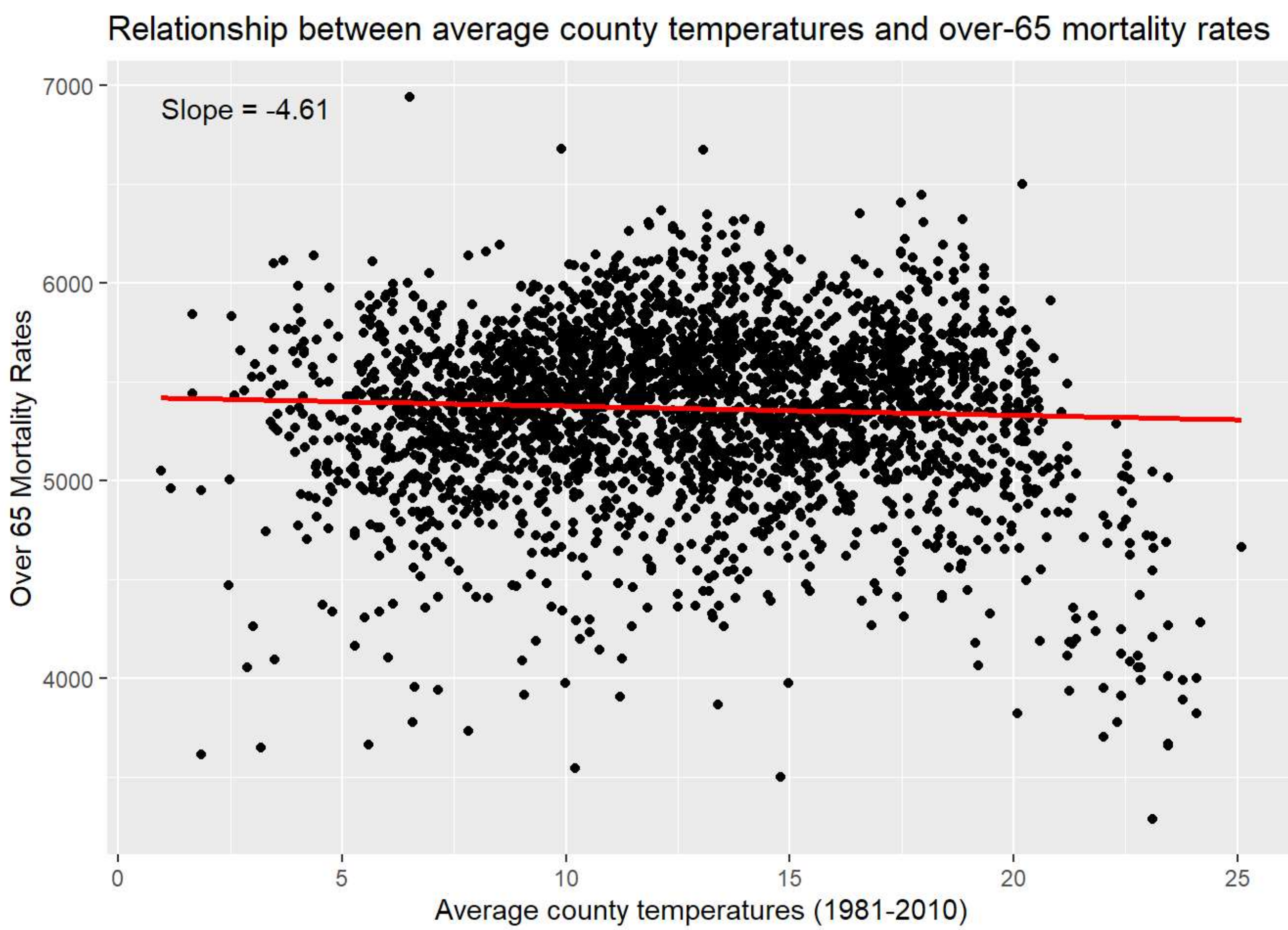
```
avg_by_county <- merge(avg_by_county, ave_temp, by = "countycode")

model <- lm(cruderate ~ normal_1981_2010, data = avg_by_county)

slope <- coef(model)[2]

ggplot(avg_by_county, aes(x = normal_1981_2010, y = cruderate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Relationship between average county temperatures and over-65 mortality rates") +
  xlab("Average county temperatures (1981-2010)") +
  ylab("Over 65 Mortality Rates") +
  annotate("text", x = min(avg_by_county$normal_1981_2010), y = max(avg_by_county$cruderate), la
bel = paste("Slope =", round(slope, 2)), hjust = 0, vjust = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



3.

b) ### Plotting graphs as instructed: hot days first, and then hotter days

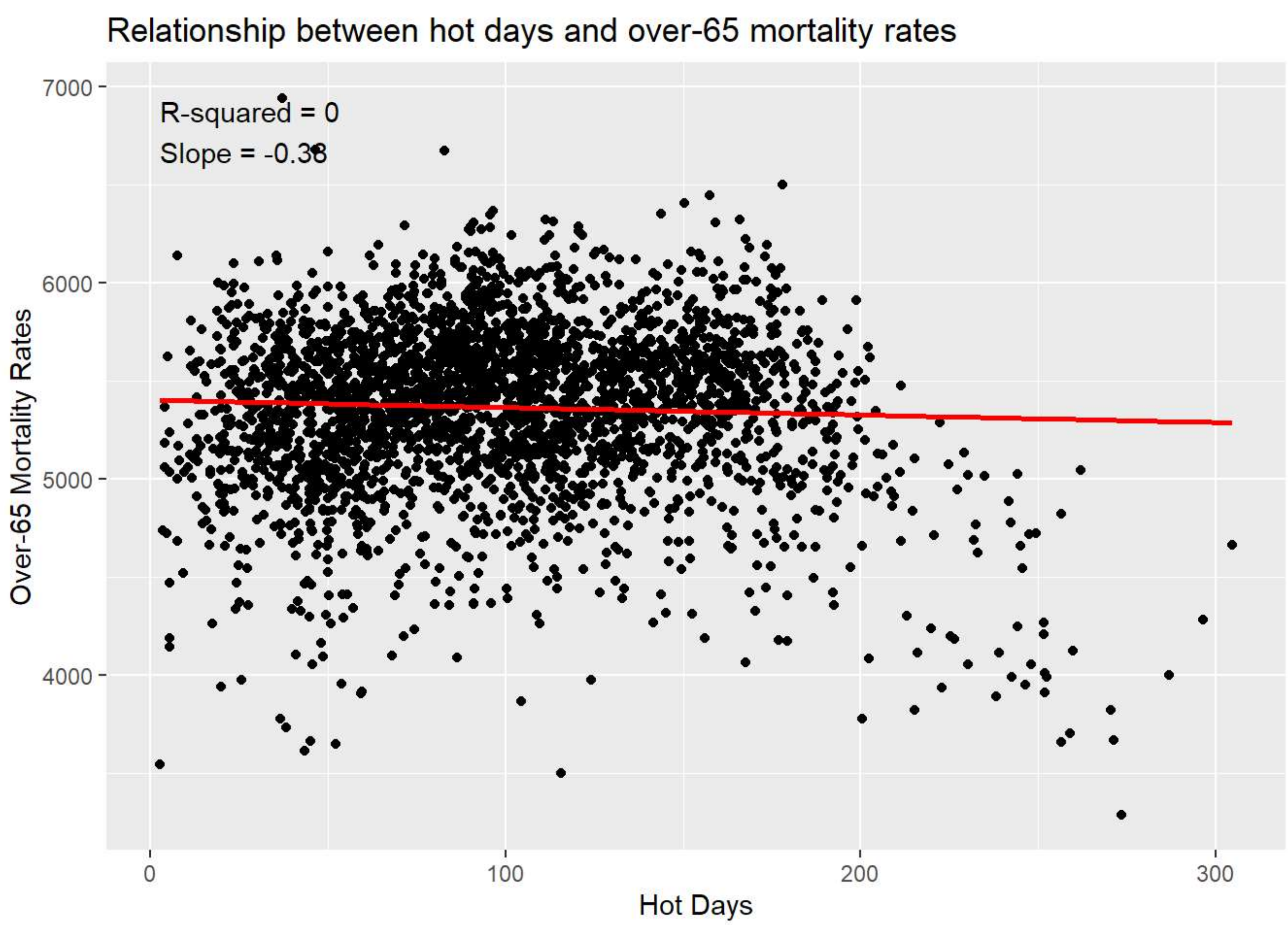
```
model_hotdays <- lm(cruderate ~ hotdays, data = avg_by_county)

slope_hotdays <- coef(model_hotdays)[2]

rsq <- summary(model_hotdays)$r.squared

ggplot(avg_by_county, aes(x = hotdays, y = cruderate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Relationship between hot days and over-65 mortality rates") +
  xlab("Hot Days") +
  ylab("Over-65 Mortality Rates") +
  annotate("text", x = min(avg_by_county$hotdays), y = max(avg_by_county$cruderate), label = pas
te("Slope =", round(slope_hotdays, 2)), hjust = 0, vjust = 3) +
  annotate("text", x = min(avg_by_county$hotdays), y = max(avg_by_county$cruderate) - 10, label
= paste("R-squared =", round(rsq, 2)), hjust = 0, vjust = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



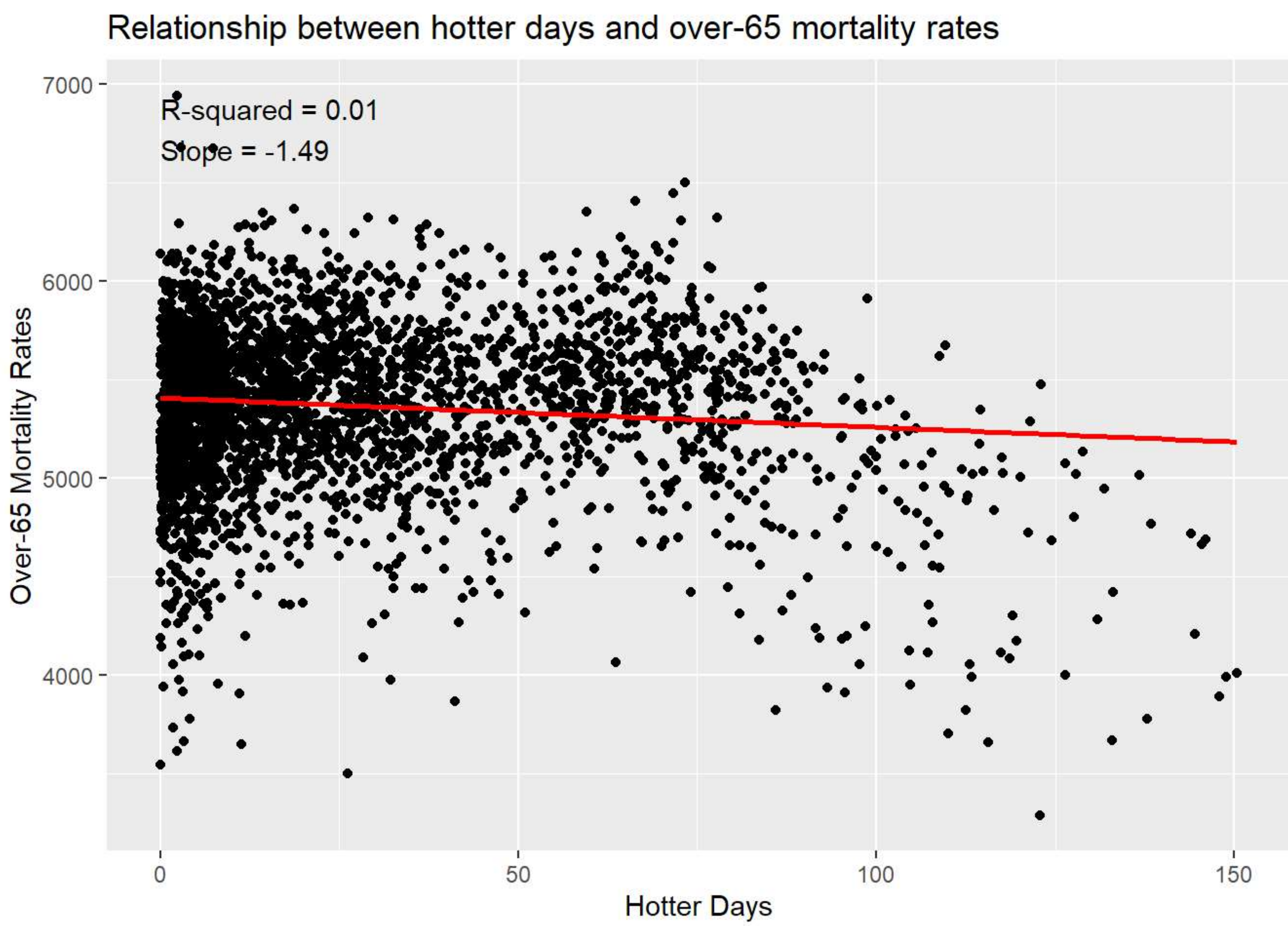
```
model_hotterdays <- lm(cruderate ~ hotterdays, data = avg_by_county)

slope_hotterdays <- coef(model_hotterdays)[2]

rsq <- summary(model_hotterdays)$r.squared

ggplot(avg_by_county, aes(x = hotterdays, y = cruderate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Relationship between hotter days and over-65 mortality rates") +
  xlab("Hotter Days") +
  ylab("Over-65 Mortality Rates") +
  annotate("text", x = min(avg_by_county$hotterdays), y = max(avg_by_county$cruderate), label =
paste("Slope =", round(slope_hotterdays, 2)), hjust = 0, vjust = 3) +
  annotate("text", x = min(avg_by_county$hotterdays), y = max(avg_by_county$cruderate) - 10, lab
el = paste("R-squared =", round(rsq, 2)), hjust = 0, vjust = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



#

Question 4 ### Data reload not needed since the dataframe “merged_df” exists consisting of the original two datasets merged together and the columns for hot days and hotter days are appended to it

```
selected_counties <- c("Mobile County, AL", "Cook County, IL", "Los Angeles County, CA", "Miami-
Dade County, FL")

# Creating new dataframe with selected counties
subset_q4 <- merged_df[merged_df$county.x %in% selected_counties, ]

# Inspecting subsetted dataframe
print(head(subset_q4))
```

```
##      countycode          county.x statecode year deaths population cruderate
## 1681       1097 Mobile County, AL         1 1972   1614      26805  6021.265
## 1682       1097 Mobile County, AL         1 1974   1616      28963  5579.533
## 1683       1097 Mobile County, AL         1 1975   1571      30099  5219.442
## 1684       1097 Mobile County, AL         1 1976   1629      31042  5247.729
## 1685       1097 Mobile County, AL         1 1973   1658      27814  5961.027
## 1686       1097 Mobile County, AL         1 1968   1431      23301  6141.367
##      tday_lt10 tday_10_20 tday_20_30 tday_30_40 tday_40_50 tday_50_60
## 1681         0          0 0.26553857   6.059636   35.08247   51.90860
## 1682         0          0 0.00744053   2.818002   24.96626   56.36346
## 1683         0          0 0.08076945   6.772363   33.28789   49.18138
## 1684         0          0 0.12955533  11.004089   41.18118   66.47477
## 1685         0          0 0.11111154  13.585549   20.14514   62.23020
## 1686         0          0 0.12316273  11.848643   53.04579   60.18242
##      tday_60_70 tday_70_80 tday_80_90 tday_gt90    cdd65    hdd65 prec_0_10
## 1681   78.76274  100.85056   92.07046         0 2728.610 1447.552         0
## 1682   93.74758  122.81201   64.28524         0 2443.549 1229.144         0
## 1683   86.17596  114.84464   74.65700         0 2555.789 1445.850         0
## 1684   75.61369  102.99716   67.59956         0 2281.359 1905.083         0
## 1685   76.67447   94.57696   97.67656         0 2801.049 1547.177         0
## 1686   55.00574   97.40452   87.38972         0 2618.101 2103.031         0
##      prec_10_15 prec_15_20 prec_20_25 prec_25_30 prec_30_35 prec_35_40
## 1681          0          0          0          0          0          0
## 1682          0          0          0          0          0          0
## 1683          0          0          0          0          0          0
## 1684          0          0          0          0          0          0
## 1685          0          0          0          0          0          0
## 1686          0          0          0          0          0          0
##      prec_40_45 prec_45_50 prec_50_55 prec_55_60 prec_gt60 division  ssyy
## 1681          0          0          1          0         0        6 11972
## 1682          0          0          0          0         1        6 11974
## 1683          0          0          0          0         1        6 11975
## 1684          0          0          0          1         0        6 11976
## 1685          0          0          0          0         1        6 11973
## 1686          1          0          0          0         0        6 11968
##           county.y state normal_1981_2010  hotdays hotterdays
## 1681 Mobile County    AL         19.53889 192.9210   92.07046
## 1682 Mobile County    AL         19.53889 187.0973   64.28524
## 1683 Mobile County    AL         19.53889 189.5016   74.65700
## 1684 Mobile County    AL         19.53889 170.5967   67.59956
## 1685 Mobile County    AL         19.53889 192.2535   97.67656
## 1686 Mobile County    AL         19.53889 184.7942   87.38972
```

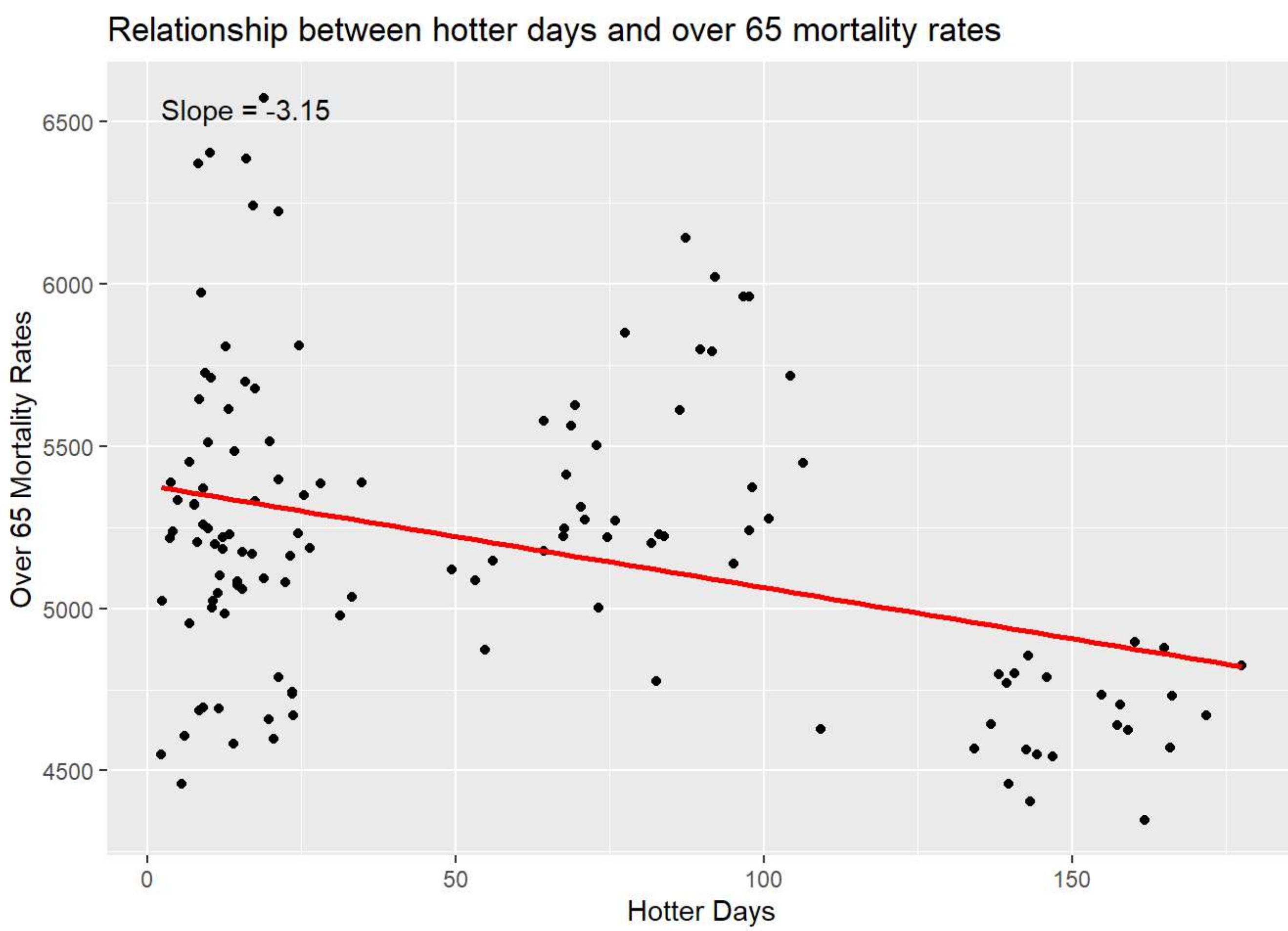
4. a)

```
model_hotterdays <- lm(cruderate ~ hotterdays, data = subset_q4)

slope_hotterdays <- coef(model_hotterdays)[2]

ggplot(subset_q4, aes(x = hotterdays, y = cruderate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("Relationship between hotter days and over 65 mortality rates") +
  xlab("Hotter Days") +
  ylab("Over 65 Mortality Rates") +
  annotate("text", x = min(subset_q4$hotterdays), y = max(subset_q4$cruderate), label = paste("S
lope =", round(slope_hotterdays, 2)), hjust = 0, vjust = 1)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

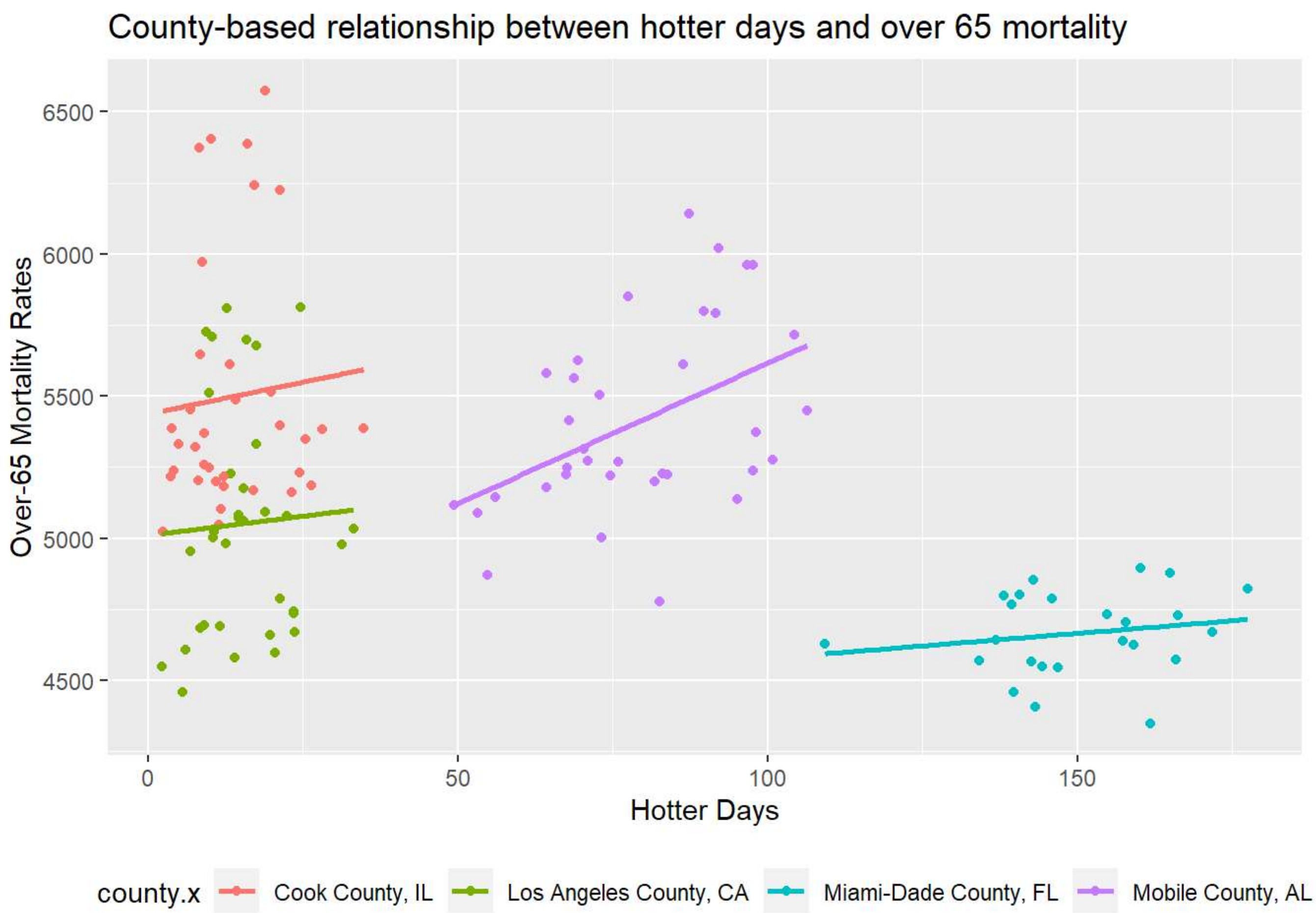


4.

b)

```r
ggplot(subset_q4, aes(x = hotterdays, y = cruderate, color = county.x)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("County-based relationship between hotter days and over 65 mortality") +
  xlab("Hotter Days") +
  ylab("Over-65 Mortality Rates") +
  theme(legend.position = "bottom")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



#

Question 5

```r
dg2011 <- felm (cruderate
                ~ tday_lt10 + tday_10_20 + tday_20_30 + tday_30_40 + tday_40_50
                + tday_50_60 + tday_70_80 + tday_80_90 + tday_gt90
                + prec_10_15 + prec_15_20 + prec_20_25 + prec_25_30 + prec_30_35
                + prec_35_40 + prec_40_45 + prec_45_50 + prec_50_55 + prec_55_60
                + prec_gt60
                | countycode + ssyy | 0 | countycode,
                data = merged_df, weight = merged_df$population)
summary(dg2011)
```

```
##
## Call:
##    felm(formula = cruderate ~ tday_lt10 + tday_10_20 + tday_20_30 +      tday_30_40 + tday_40
_50 + tday_50_60 + tday_70_80 + tday_80_90 +      tday_gt90 + prec_10_15 + prec_15_20 + prec_20_
25 + prec_25_30 +      prec_30_35 + prec_35_40 + prec_40_45 + prec_45_50 + prec_50_55 +      pre
c_55_60 + prec_gt60 | countycode + ssyy | 0 | countycode,      data = merged_df, weights = merge
d_df$population)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -304609  -18444    -245   18621  279903
##
## Coefficients:
##            Estimate Cluster s.e. t value Pr(>|t|)
## tday_lt10    3.7239       1.3839   2.691 0.007167 **
## tday_10_20   2.6755       1.1061   2.419 0.015625 *
## tday_20_30   3.5811       0.8509   4.209 2.64e-05 ***
## tday_30_40   1.8146       0.6684   2.715 0.006669 **
## tday_40_50   1.4530       0.5699   2.550 0.010833 *
## tday_50_60   0.2559       0.3271   0.782 0.434137
## tday_70_80   0.8364       0.6805   1.229 0.219131
## tday_80_90   1.3660       0.8861   1.542 0.123280
## tday_gt90    5.3466       1.3828   3.867 0.000113 ***
## prec_10_15  -1.5165      20.9606  -0.072 0.942330
## prec_15_20 -10.2076      30.6655  -0.333 0.739256
## prec_20_25 -17.3908      36.4208  -0.477 0.633044
## prec_25_30 -12.1570      39.9516  -0.304 0.760925
## prec_30_35 -15.2506      42.5402  -0.358 0.719995
## prec_35_40  -5.2618      45.0634  -0.117 0.907054
## prec_40_45  -3.1903      46.0152  -0.069 0.944730
## prec_45_50  -9.4240      47.0961  -0.200 0.841415
## prec_50_55   2.7343      47.6148   0.057 0.954210
## prec_55_60  -6.3052      48.3366  -0.130 0.896225
## prec_gt60   -5.8340      49.6347  -0.118 0.906440
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30010 on 99838 degrees of freedom
## Multiple R-squared(full model): 0.7612   Adjusted R-squared:  0.75
## Multiple R-squared(proj model): 0.001653   Adjusted R-squared: -0.04506
## F-statistic(full model, *iid*):68.14 on 4671 and 99838 DF, p-value: < 2.2e-16
## F-statistic(proj model): 2.187 on 20 and 2985 DF, p-value: 0.001731
```

5. a)

```
# Creating a dataframe with order of temperature bins, with the value of zero for the temperature bin of 60-70F
tempbins <- data.frame(
  temperature_bins = factor(c("<10", "10-20", "20-30", "30-40", "40-50", "50-60", "60-70", "70-80", "80-90", ">90"),
                            levels = c("<10", "10-20", "20-30", "30-40", "40-50", "50-60", "60-70", "70-80", "80-90", ">90")),
  estimates = c(3.7239, 2.6755, 3.5811, 1.8146, 1.4530, 0.2559, 0, 0.8364, 1.3660, 5.3466),
  std_errors = c(1.3839, 1.1061, 0.8509, 0.6684, 0.5699, 0.3271, 0, 0.6805, 0.8861, 1.3828)
)

ggplot(tempbins, aes(x = temperature_bins, y = estimates)) +
  geom_point() +
  geom_errorbar(aes(ymin = estimates - 2*std_errors, ymax = estimates + 2*std_errors), width = 0.2) +
  geom_line(aes(group = 1)) +
  geom_text(aes(label = estimates), vjust = -1.5, hjust = 0.5) +
  labs(x = "Temperature bins (F)",
       y = "Over-65 mortality rate",
       title = "Estimated impact of a day in 9 daily mean temperature (F) bins on\nannual over-65 mortality rate, relative to a day in the 60-70F bin\n(including +2 and -2 standard errors)")
```