7/6/2019 liana\_code.R

```
1
2
   3
  4
   setwd("/Users/joanmeiners/Dropbox/Fall 2017/Environmental
   Journalism/Liana Solar EJUF/")
6
7
   library(dplyr)
8
   library(ggplot2)
9
10 # Load data
   solar = read.csv("Liana Solar.csv")
11
12 | solar$kWDC = as.numeric(solar$kWDC) # make sure kilowatt-hours reading in as numeric
   solar$ADDRESS = trimws(solar$ADDRESS) # delete extra whitespace around addresses
13
   dim(solar)
14
15 View(solar)
16
17 | # join to value dataset and pull in zip code information from community parcels
   dataset
18 value = read.csv("/Users/joanmeiners/Dropbox/Fall 2017/Environmental
   Journalism/value.csv", header = TRUE)
19 | solarval = plyr::join(solar, value, by = "ADDRESS")
20 | solarval = subset(solarval, POSTAL != "NA" & CNTASSDVALUE < 20000000, select =
   c("ADDRESS", "POSTAL", "kWDC", "CNTASSDVALUE")) # filter out all addresses without an
   associated zipcode and all properties valued at over $2 million since they are likely
   not personal residences
21 solarval$POSTAL = as.factor(solarval$POSTAL)
22 solarval$ADDRESS = as.character(solarval$ADDRESS)
23
24 # transform variables to reasonable degree of normalcy
25
   hist(solarval$kWDC) # needs to be transformed
   hist(log10(solarval$kWDC)) # roughly normally distributed, not perfect (right skewed)
26
   hist(log10(solarval$CNTASSDVALUE)) # pretty normal, slightly left skewed
27
28
29
   # look for trends in solar usage by property value
   solar1 = glm(log10(kWDC) ~ log10(CNTASSDVALUE), data = solarval)
30
   summary(solar1)
31
32
   solar1
33
34 # graph the results
35
   quartz(width = 12, height = 6)
   ggplot(aes(y = kWDC, x = CNTASSDVALUE), data = solarval) +
     scale x log10(breaks = c(2000000, 200000, 20000), limits = c(19000, 3000000),
37
   labels = function(x) paste0("$", scales::comma(x))) +
     scale_y = c(0, 25, 50, 100, 200, 400) +
38
39
     geom point(color = "grey") +
     xlab("County-assessed Property Value (USD)") + ylab("Solar energy usage per address
40
   (kWDC)") +
41
     theme(axis.title = element text(family = "Trebuchet MS", color="#666666",
   face="bold", size=15)) +
     theme(axis.text = element_text(family = "Trebuchet MS", color="#666666",
42
   face="bold", size=12)) +
     geom_smooth(method = "lm", se=FALSE, color="darkgreen")
43
44
45
   # save figure to file
```

```
tiff(filename = "Solar_value.tiff", units = "in", compression = "lzw", res = 300,
width = 12, height = 6)
# run figure code from above section here (w/o quartz line), then next step
dev.off()
```