```r
1   ################################################################
2   ############# EJUF Code to Evaluate Solar Data for Liana #############
3   ####################### Joan Meiners 2017 #######################
4
5   setwd("/Users/joanmeiners/Dropbox/Fall 2017/Environmental
    Journalism/Liana_Solar_EJUF/")
6
7   library(dplyr)
8   library(ggplot2)
9
10  # load data
11  solar = read.csv("Liana_Solar.csv")
12  solar$kWDC = as.numeric(solar$kWDC) # make sure kilowatt-hours reading in as numeric
13  solar$ADDRESS = trimws(solar$ADDRESS) # delete extra whitespace around addresses
14  dim(solar)
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 < 2000000, 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
26  hist(log10(solarval$kWDC)) # roughly normally distributed, not perfect (right skewed)
27  hist(log10(solarval$CNTASSDVALUE)) # pretty normal, slightly left skewed
28
29  # look for trends in solar usage by property value
30  solar1 = glm(log10(kWDC) ~ log10(CNTASSDVALUE), data = solarval)
31  summary(solar1)
32  solar1
33
34  # graph the results
35  quartz(width = 12, height = 6)
36  ggplot(aes(y = kWDC, x = CNTASSDVALUE), data = solarval) +
37    scale_x_log10(breaks = c(2000000 ,200000, 20000), limits = c(19000, 3000000),
    labels = function(x) paste0("$", scales::comma(x))) +
38    scale_y_log10(breaks = c(0, 25, 50, 100, 200, 400)) +
39    geom_point(color = "grey") +
40    xlab("County-assessed Property Value (USD)") + ylab("Solar energy usage per address
    (kWDC)") +
41    theme(axis.title = element_text(family = "Trebuchet MS", color="#666666",
    face="bold", size=15)) +
42    theme(axis.text = element_text(family = "Trebuchet MS", color="#666666",
    face="bold", size=12)) +
43    geom_smooth(method = "lm", se=FALSE, color="darkgreen")
44
45  # save figure to file
```

```r
46  tiff(filename = "Solar_value.tiff", units = "in", compression = "lzw", res = 300,
    width = 12, height = 6)
47  # run figure code from above section here (w/o quartz line), then next step
48  dev.off()
49
```