Challenges in the Visualization of Sleep Deprivation across the US

Michael Geden

Abstract

Sleep deprivation is a wide-spread public health issue in the United States and many other countries. Rising awareness of the issue has led to large-scale epidemiological efforts to measure the prevalance and cost of the issue. Visualization of this information is crucial for informing policy and planning targeted interventions to populations in need, however, visualizing the data often involves a large number of decisions on the preprocessing of the data which can influence the conclusions made. This brief report touches on a few of these challenges, and provides some recommendations for transparency.

Introduction

Approximately 1/3 of the population of the United States reports having less sleep on average than the recommended 7 hours a night (CDC, 2011; Liu, 2016). Chronic sleep deprivation is considered a serious public health issue, as insufficient sleep is associated with increased workplace related accidents (Dinges, 1995; Rosekind et al., 2010), obesity (Gangwisch, Malaspina, Boden-Albala, & Heymsfield, 2004; Knutson, Spiegel, Penev, & Van Cauter, 2004), drowsy driving (J. Horne & Louise, 1995; Howard et al., 2004), cardiovascular disease (Ayas et al., 2003; Mullington, Hacch, Toth, Serrador, & Meier-Ewert, 2009), and a variety of other risks/conditions. Sleep deprivation is also correlated with decreased decision making (Harrison & Horne, 2000) and sustained attention (Lim & Dinges, 2000), which may provide one explanation for some of the associated behavioral detriments.

There have been sizable efforts on the part of the Centers for Disease Control (CDC) and National Sleep Foundation (NSF) to better understand the prevalance and costs of sleep deprivation through large-scale surveys across the United States. These surveys provide critical information for future policy and planning targeted interventions, and have led to some preliminary suggestions including employers tailoring shift-systems design (CDC, 2012), greater involvement by health-care providers (CDC, 2009), limiting active technology use before sleep (Gradisar et al., 2013), and general increased public awareness of sleep deprivation.

Communicating complex spatial data, such as nation-wide polls, requires a number of decisions and pre-processing steps before creating the final visualizations. This preliminary work can have large effects on the final product, leading to potentially different conclusions. The purpose of this brief report is to outline some practices which would help improve the transparency and clarity of the final report employing these visualizations.

Data

For this brief report we will the 2009 Behavioral Risk Factor Surveillance System (BRFSS) results collected by the CDC. The BRFSS is a yearly survey intended to measure health related behaviors and demographics such as tobacco use, seatbelt use, sleep related habits, etc. The 2009 BRFSS was collected through landlines only, with cellphones being added in 2011 to the survey sample. A total of 432,607 records were collected in this data set across 2231 of the total 3109 counties in the contiguous United States. The sleep related question that will be used here presented below.

- "During the past 30 days, for about how many days have you felt you did not get enough rest or sleep? (number of days)"

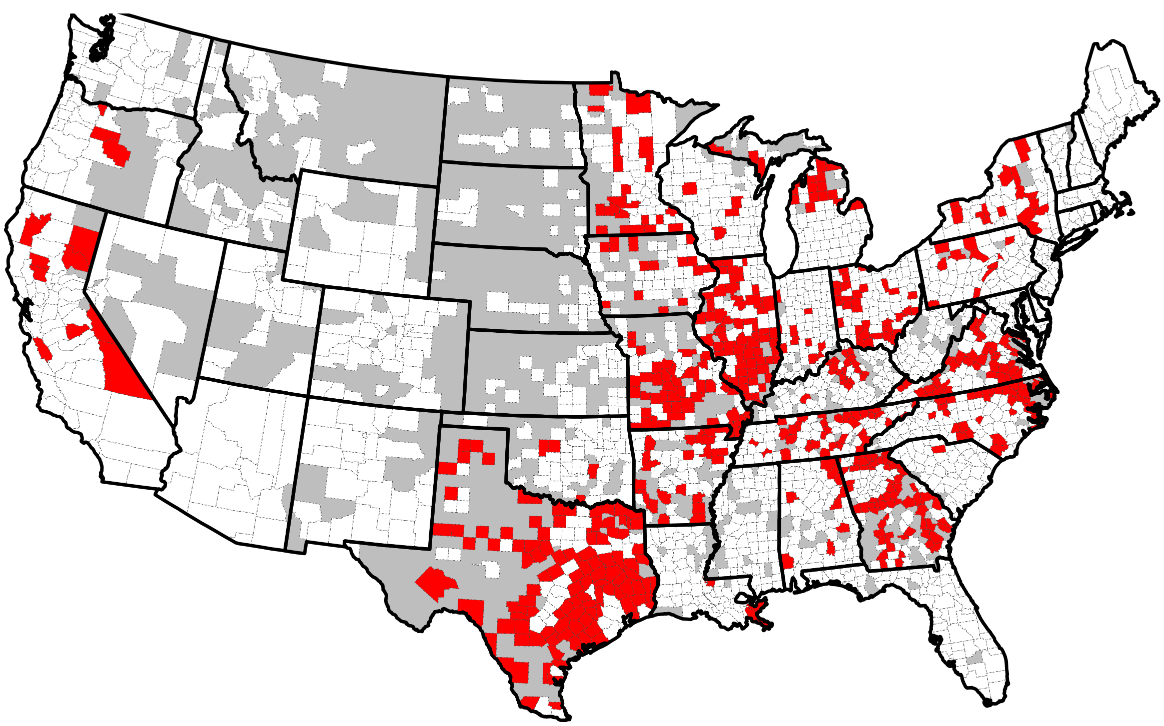
This dataset and question were selected due to their use in Grandner et al. (2015), which will allow for some conrete discussion points and comparisons. The responses to this question were dichotomized, with participants reporting greater than 14 days of insufficient sleep being sleep deprived. This cutoff was chosen based on previous research using the cutoff (Grandner et al., 2015; Strine & Chapman, 2015) and some support found for the practical significance of it in Edinger et al. (2011).

Challenges

Sample Size

(Grandner et al., 2015) Didn't discuss this, and many of the regions with high quintile were also the regions in which there was limited data

Frequency of Responses

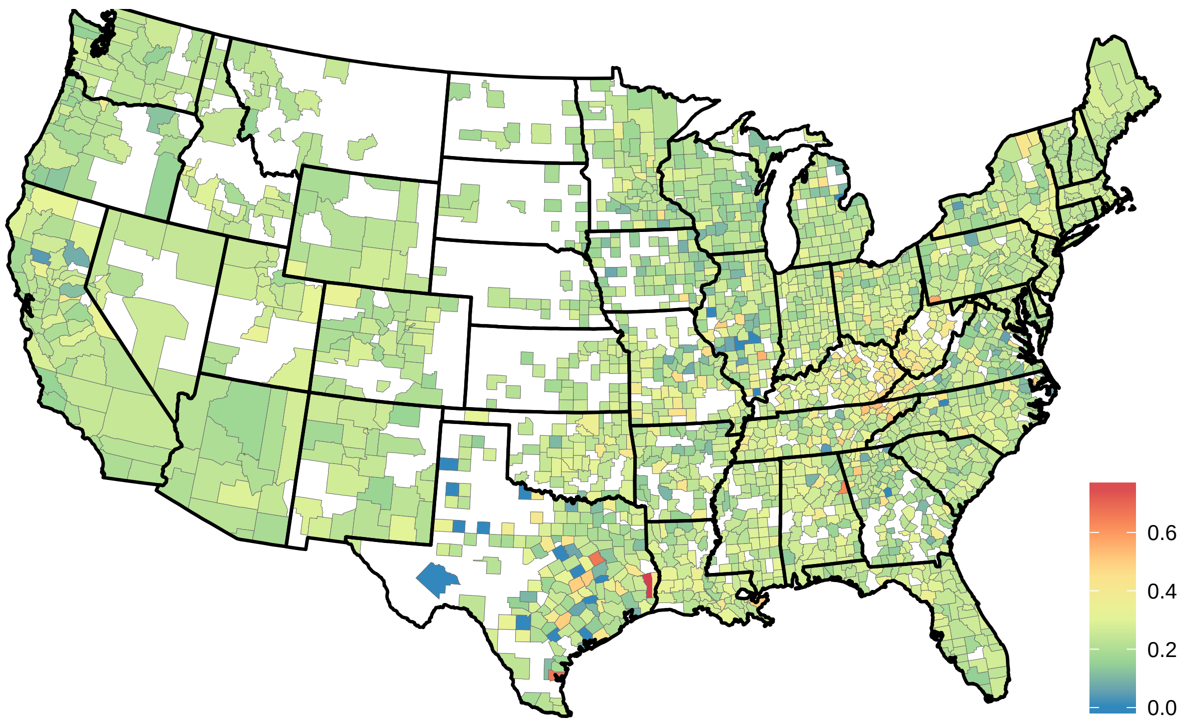


Missing Data

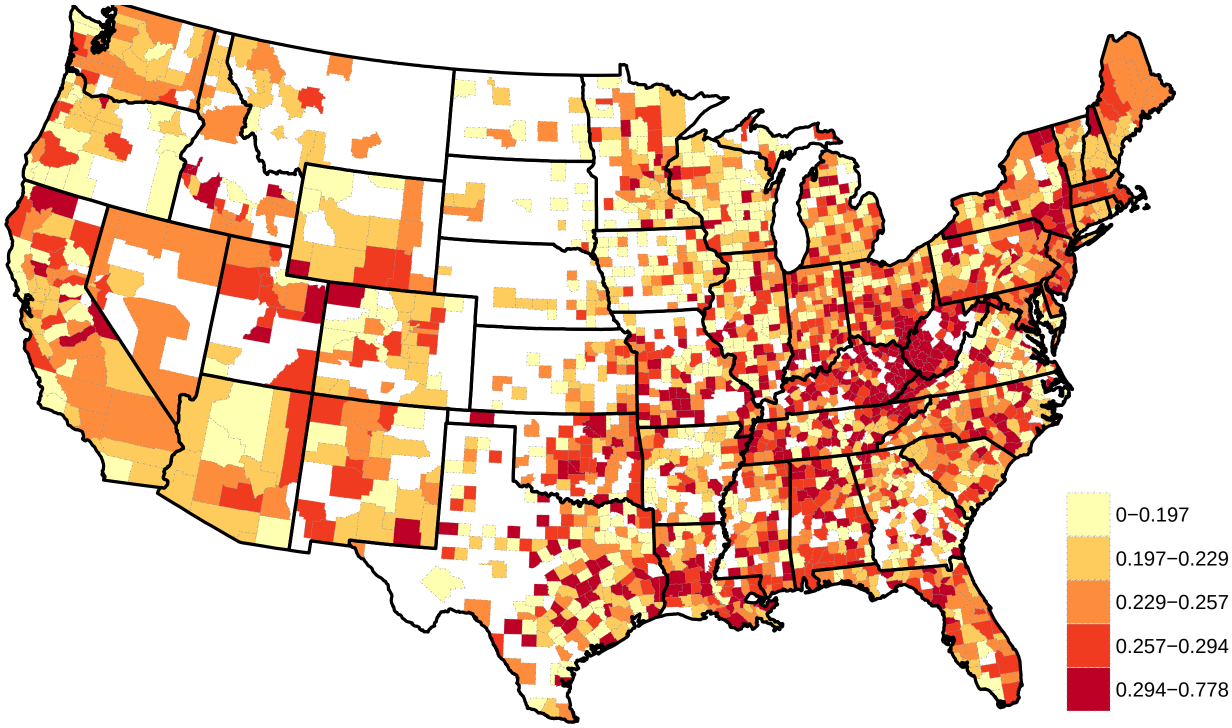
Important to address, and quite hard to. It would be good at the very least to look at the proportion of the missing data (excluding refusal) per county and the proportion of refused to respond per county

Quintiles

Continuous Proportion of Sleep Deprivation



Quintiles of Proportion of Sleep Deprivation



Adjusted Proportions

THIS NEEDS A LOT MORE IF GOIGN TO INCLUDE. MIGHT NOT HAVE THE TIME TO GO HAM ENOUGH

State level could use asmyptotic normal confidence intervals, but the smaller sample sizes in the county level analysis would need to use an exact test or bootstrap estimate (provided moderate n, ~15-30). Given the presence of numerous other variables not controlled for, and the differences in sampling error for older populations vs. younger populations, simply using age-adjusted estimates would likely not be sufficients. Instead using a multilevel approach such as the one outlined in Zhang et al. (2014)

Transparency

All information used in creating the figures and analysis should be provided somewhere as well. For example, here I have used the same data and methodology as (Grandner et al., 2015) for creating the quintile plot, yet have gotten a few sizable differences. This could be due to a variety of reasons, however without being able to compare code it is unclear whether this was due to a mistake on one of our parts, or a different decision in the preprocessing of the data that wasn't clearly described.

Conclusion

Code

```
BRFSS_geoextraction <- function(filepath, id = "county", position,
                                extraction, contiguous=TRUE,latlong=TRUE,
                                func = function(x) {mean(x, na.rm=T)}) {
    # Error Catching
    if(!grepl("state", tolower(position[1]))){stop("State must be first position.")}
    if(!is.character(filepath)) {stop("Filepath must be character")}
    if(!is.character(position)) {stop("Position must be character")}
    if(!is.character(extraction)) {stop("Extraction variable must be character name.")}
    if(length(extraction)!=1) {stop("Function currently only accepts one extracted var.")}
    if(!id %in% c("county", "state")) {stop("ID must be county or state.")}
    if(!is.function(func)){stop("Func must be a function.")}
    if(id == "county" & length(position) != 2) {
        stop("County level requires position to have State and County varialbe names. ")
    }
    # Setup
    require(SASxport)
    require(dplyr)
    require(maps)
    require(stringr)
    require(ggplot2)
    strsplit2 <- function(x, char, index){
        unlist(strsplit(x, char))[index]
    }
    clear_labels <- function(x) {
        if(is.list(x)) {
            for(i in 1 : length(x)) class(x[[i]]) <- setdiff(class(x[[i]]), 'labelled')
            for(i in 1 : length(x)) attr(x[[i]],"label") <- NULL
        }
        else {
            class(x) <- setdiff(class(x), "labelled")
            attr(x, "label") <- NULL
        }
        return(x)
    }
    # Read Data
    data <- read.xport(filepath) %>%
        clear_labels(.)
    # Subset
    keep <- c(position, extraction)
    keep.states <- if(contiguous) {
        c(15, 2, 66, 72, 78)
    } else {c()}
    # Remove Excess
    data2 <- data %>%
        select(one_of(keep)) %>% # Select columns of interest
```

```r
    filter(!(!!rlang::sym(position[1])) %in% keep.states) %>% # Select States
    filter(!(is.na(!!rlang::sym(position[1])))) # remove missing states
if(id == "county"){
    data2 <- data2 %>%
        filter(!(!!rlang::sym(position[2])) %in% c(777,999)) %>%
        filter(!(is.na(!!rlang::sym(position[2])))) %>%
        mutate(FIPS = as.numeric(paste0(
            !!rlang::sym(position[1]),
            str_pad(!!rlang::sym(position[2]),3,pad = "0"),
            sep = "")
        ))
} else {
    data2 <- data2 %>%
        mutate(FIPS = !!rlang::sym(position[1]))
}
# Get FIPS Codes
if(id == "county") {
    data(county.fips)
    county.fips$polyname <- sapply(county.fips$polyname, function(x) {
        unlist(strsplit(x, ":"))[1]
    })
    county.fips <- unique(county.fips)
    fips.codes <- county.fips
} else {
    data(state.fips)
    state.fips$polyname <- sapply(state.fips$polyname, function(x) {
        unlist(strsplit(x, ":"))[1]
    })
    state.fips <- unique(state.fips)
    fips.codes <- as.data.frame(state.fips)
}
varname <- paste0(extraction, ".f", sep = "")
# Merge By FIPS
data3 <- data2 %>%
    left_join(fips.codes, by = c("FIPS" = "fips")) %>% # Get location string
    group_by(polyname,FIPS) %>% # State/County-wise operations
    summarize(!!varname := func(!!rlang::sym(extraction)), # Function
              n = sum(!is.na(!!rlang::sym(extraction))), # Frequency of response
              prop.responded = sum(!is.na(!!rlang::sym(extraction)))/n()) %>%
    filter(!is.na(!!rlang::sym(varname))) %>% # Remove Missing
    mutate(state = strsplit2(polyname, ",", 1)) # Extract state name
if(id == "county") {
    data3$county <- strsplit2(data3$polyname, ",",2)
}
# Merge for coordinates
if(id == "county"){
    map <- map_data('county') %>%
```

```r
            mutate(polyname = paste0(region, ",", subregion, sep = "")) %>%
            select(-c(group, order, region, subregion))
    } else {
        map <- map_data('state') %>%
            mutate(polyname = paste0(region, ",", subregion, sep = "")) %>%
            select(-c(group, order, region, subregion))
    }
    if(latlong) {data3 <- left_join(data3, map, by = "polyname")}
    # Output
    data3
}
```

References

Ayas, N., White, D., Manson, J., Stampfer, M., Speizer, F., Malhotra, A., & Hu, F. (2003). Archives of internal medicine. *Progress in Cardiovascular Diseases*, *2*(163), 205–2009.

CDC. (2009). Perceived insufficient rest or sleep among adults - united states, 2008. *Morbidity and Mortality Weekly Report*, *42*(58), 1175–1179.

CDC. (2011). Effect of short sleep duration on daily activities - united states, 2005-2008. *Morbidity and Mortality Weekly Report*, *8*(60), 239.

CDC. (2012). Short sleep duration among workers - united states, 2010. *Morbidity and Mortality Weekly Report*, *16*(61), 281–285.

Dinges, D. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, *14*(4), 4–14.

Edinger, J., Wyatt, J., Stepanski, E., Olsen, M., Stechuchak, K., Carney, C., ... Krystal, A. (2011). Testing the reliability and validity of dsm-iv-tr and icsd-2 insomnia diagnoses: Results of a multitrait-multimethod analysis. *Arch Gen Psychiatry*, *10*(68), 992–1002.

Gangwisch, J., Malaspina, D., Boden-Albala, B., & Heymsfield, S. (2004). Inadequate sleep as a risk factor for obesity; analysis of the nhanes 1. *Sleep*, *10*(28), 1289–1296.

Gradisar, M., Wolfson, A., Harvey, A., Hlae, L., Rosenberg, R., & Czeisier, C. (2013). The sleep and technology use of americans: Findings from the national sleep foundation's 2011 sleep in america poll. *Journal of CLinical Sleep Medicine*, *12*(9), 1291–1299.

Grandner, M., Smith, M., Jackson, N., Jackson, T., Burgard, S., & Branas, C. (2015). Geographic distribution of unsufficient sleep across the united state: A county-level hotspot analysis. *Sleep Health*, *3*(1), 158–165.

Harrison, Y., & Horne, J. (2000). The impact of sleep deprivatino on decision making: A review. *Journal of Experimental Psychology: Applied*, *6*(3), 236–249.

Horne, J., & Louise, R. (1995). Sleep related vehicle accidents. *British Medical Journal*, *6979*(310), 565–567.

Howard, M., Desai, A., Grunstein, R., Hukins, C., Armstrong, J., Joffe, D., ... Pierce, R. (2004). Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. *American Journal of Respiratory and Critical Care Medicine*, *9*(170), 1014–1021.

Knutson, K., Spiegel, K., Penev, P., & Van Cauter, E. (2004). The metabolic consequences of sleep deprivation. *Sleep Medicine Reviews*, *3*(11), 163–178.

Lim, J., & Dinges, D. (2000). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1*(1129), 305–322.

Liu, Y. (2016). Prevalence of healthy sleep duration among adults- united states, 2014. *Morbidity and Mortality Weekly Report*, (65), 1025–1033.

Mullington, J., Hacch, M., Toth, M., Serrador, J., & Meier-Ewert, H. (2009). Progress in cardiovascular diseases. *Progress in Cardiovascular Diseases*, *4*(51), 294–302.

Rosekind, M., Gregoy, K., Mallis, Melissa, Brandt, S., Seal, B., & Lerner, D. (2010). An overview of sleepiness and accidents. *American College of Occupational and Environmental Medicine*, *52*(1), 91–98.

Strine, T., & Chapman, D. (2015). Associations of frequent sleep insufficiency with health-related quality of life and health behaviors. *Sleep Medicine*, *1* (6), 23–27.

Zhang, X., Holt, J., Lu, H., Wheaton, A., Ford, Earl, Greenlund, K., & Croft, J. (2014). Mulilevel regression and poststratification for small-area estimation of population health outcomes: A case study of crhonic obstructive pulmonary disease prevelance using the behavioral risk factor surveillance system. *American Journal of Epidemiology*, *179* (8), 1025–1033.