Health Risk and Worker Density

Nathan Seto

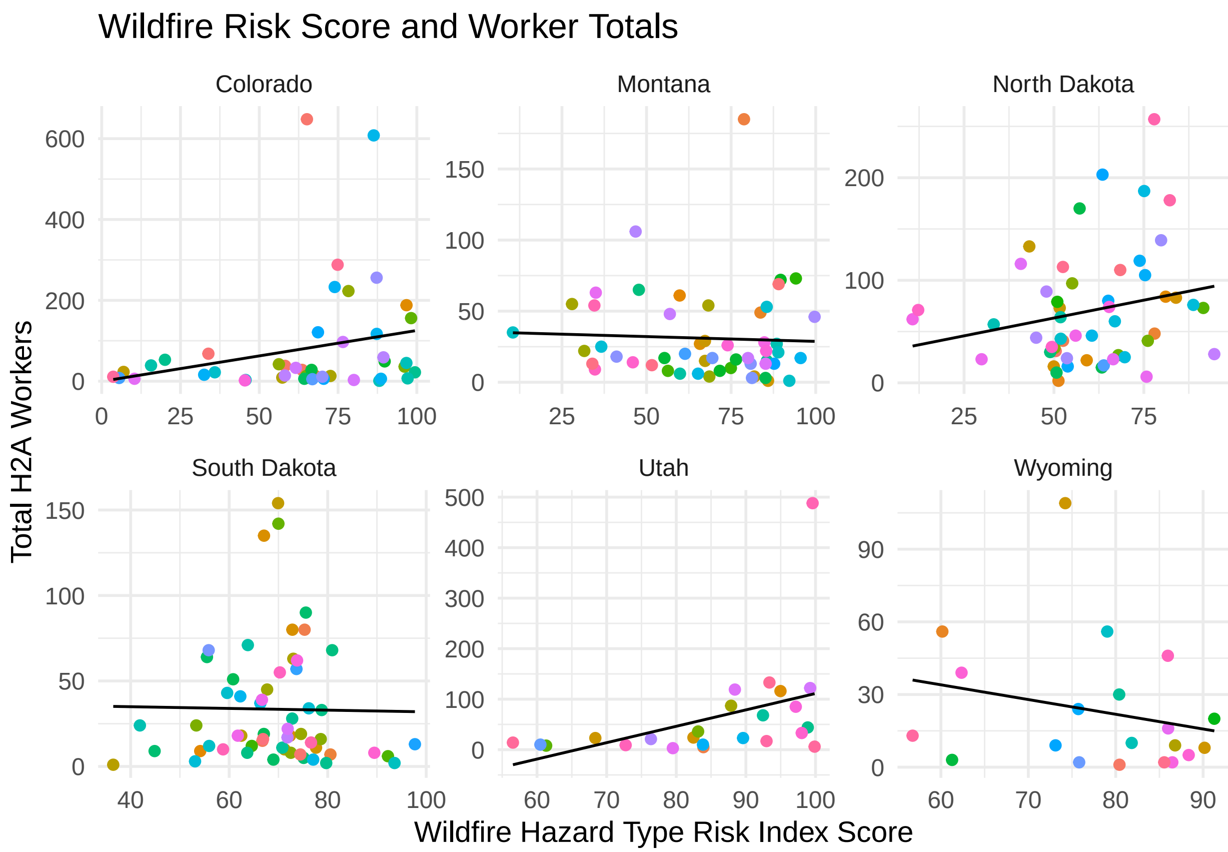
2025-02-10

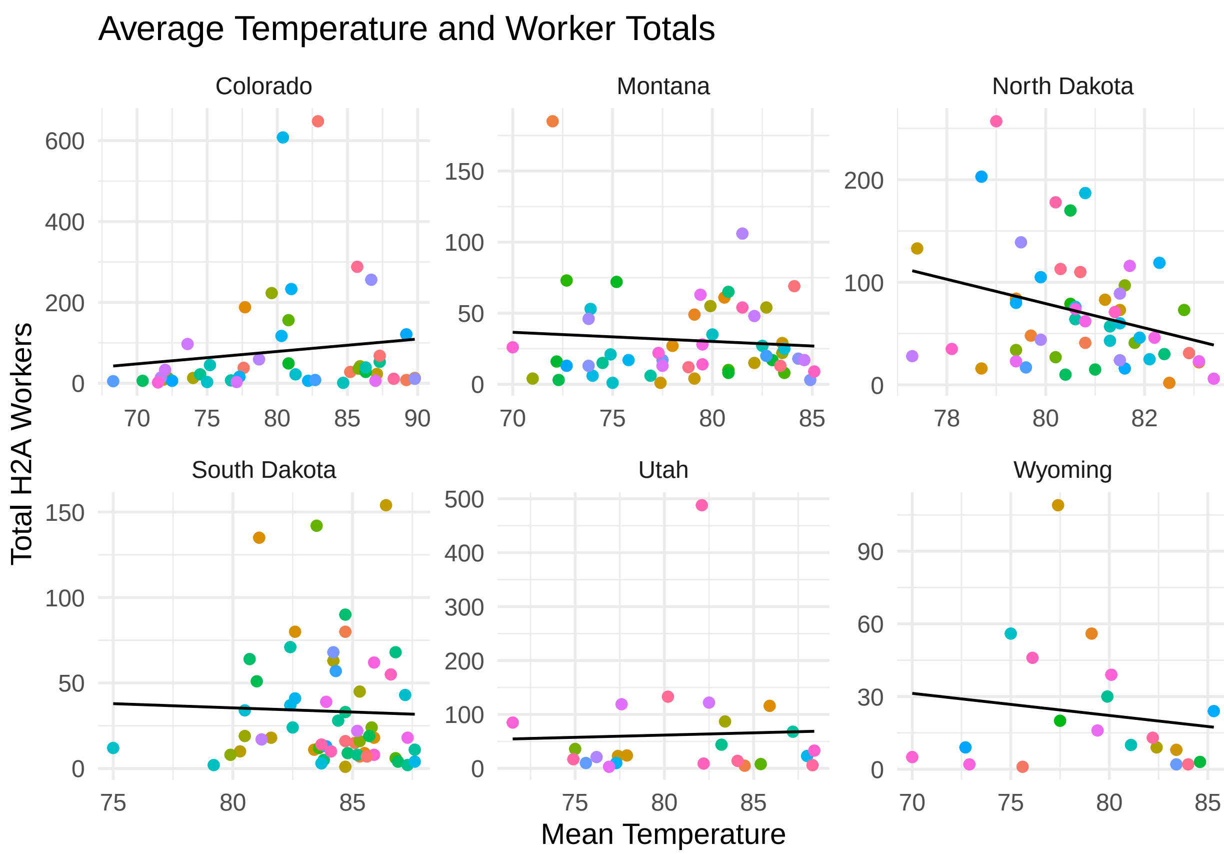
(1) What is the relationship between health risk factors and worker densities in the HICAHS area?

i. calculate unconditional correlations between worker concentrations (e.g., totals) and fire and heat environmental exposure scales (identified on website as capturing aspects of "human risk")

ii.similar correlations for other available indicators that may be of interest

iii. use county-level as the geo-observational unit





State	Multiple R-Sqr	Pearson Corr. Coef.
Colorado	0.017	0.130
Wyoming	0.019	-0.138
North Dakota	0.077	-0.277
South Dakota	0.008	-0.089
Utah	0.000	0.010

```
## Linear Model of Maximum Temperature and Worker Totals:

##
## Summary Statistics
## ======================================================
##                              Dependent variable:
##                          ----------------------------
##                                 total_workers
## ------------------------------------------------------
## wildfire_exposure_total           0.000***
##                                    (0.000)
##
## Constant                          41.902***
##                                    (5.085)
##
## ------------------------------------------------------
## Observations                         252
## R2                                  0.080
## Adjusted R2                         0.076
```

```
## Residual Std. Error          74.599 (df = 250)
## F Statistic              21.704*** (df = 1; 250)
## ===================================================
## Note:                   *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Linear Model of Wild Fire Hazard Index Score and Worker Totals:
```

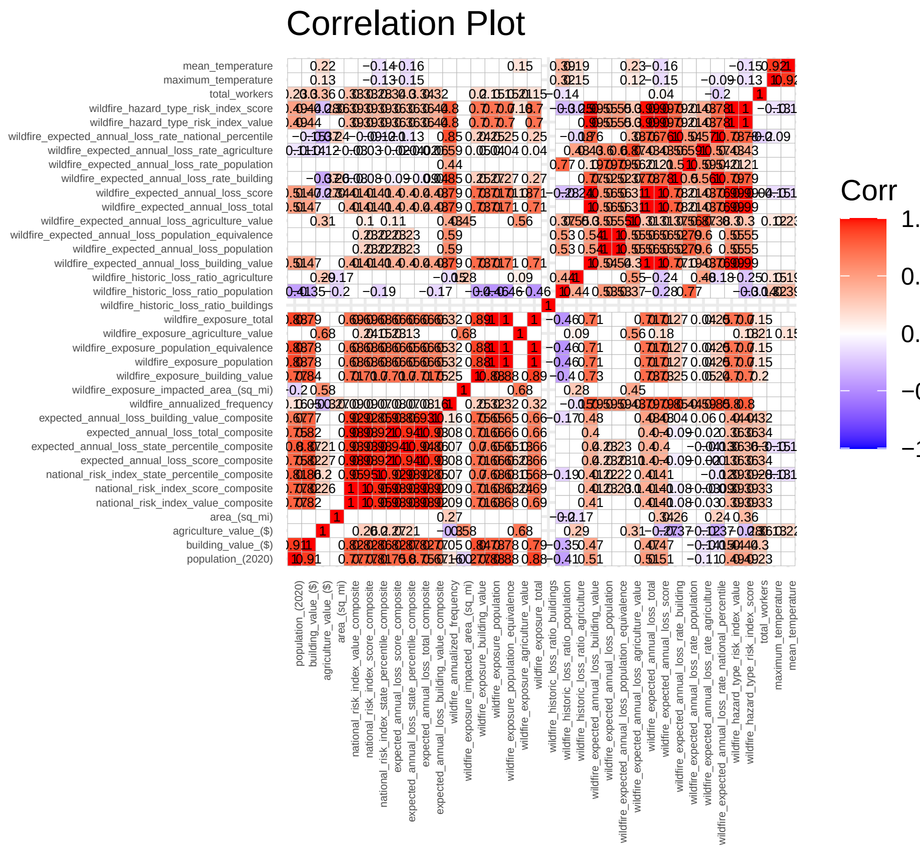
```
##
## Summary Statistics of Coefficients
## ===================================================================
##                                           Dependent variable:
##                                       ---------------------------
##                                              total_workers
## -------------------------------------------------------------------
## wildfire_hazard_type_risk_index_score          0.476*
##                                                (0.246)
##
## Constant                                       18.508
##                                               (17.431)
##
## -------------------------------------------------------------------
## Observations                                     252
## R2                                              0.015
## Adjusted R2                                     0.011
## Residual Std. Error                        77.192 (df = 250)
## F Statistic                              3.757* (df = 1; 250)
## ===================================================================
## Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Linear Model of Fire Exposure Total and Worker Totals:
```

```
##
## Summary Statistics of Coefficients
## ===================================================
##                           Dependent variable:
##                       ---------------------------
##                              total_workers
## ---------------------------------------------------
## wildfire_exposure_total          0.000***
##                                   (0.000)
##
## Constant                         41.902***
##                                   (5.085)
##
## ---------------------------------------------------
## Observations                        252
## R2                                 0.080
## Adjusted R2                        0.076
## Residual Std. Error          74.599 (df = 250)
## F Statistic              21.704*** (df = 1; 250)
## ===================================================
## Note:                   *p<0.1; **p<0.05; ***p<0.01
```

```
##
```

```
##
##  Correlation of all Variables with Total Workers
## # A tibble: 35 x 2
##    column                                                   correlation
##    <chr>                                                          <dbl>
##  1 agriculture_value_($)                                        0.359
##  2 expected_annual_loss_score_composite                         0.343
##  3 expected_annual_loss_total_composite                         0.343
##  4 national_risk_index_value_composite                          0.325
##  5 national_risk_index_score_composite                          0.325
##  6 expected_annual_loss_building_value_composite                0.324
##  7 wildfire_exposure_impacted_area_(sq_mi)                      0.302
##  8 building_value_($)                                           0.299
##  9 expected_annual_loss_state_percentile_composite              0.298
## 10 national_risk_index_state_percentile_composite               0.281
## 11 population_(2020)                                            0.234
## 12 wildfire_exposure_agriculture_value                          0.213
## 13 wildfire_exposure_building_value                             0.201
## 14 wildfire_exposure_total                                      0.149
## 15 wildfire_exposure_population                                 0.146
## 16 wildfire_exposure_population_equivalence                     0.146
## 17 wildfire_historic_loss_ratio_agriculture                     0.0661
## 18 wildfire_expected_annual_loss_agriculture_value              0.0598
## 19 wildfire_expected_annual_loss_building_value                 0.0434
## 20 wildfire_expected_annual_loss_total                          0.0424
## 21 wildfire_expected_annual_loss_score                          0.0424
## 22 wildfire_hazard_type_risk_index_value                        0.0208
## 23 wildfire_hazard_type_risk_index_score                        0.0208
## 24 area_(sq_mi)                                                 0.0139
## 25 mean_temperature                                            -0.00785
## 26 maximum_temperature                                         -0.0584
## 27 wildfire_expected_annual_loss_population                    -0.0711
## 28 wildfire_expected_annual_loss_population_equivalence        -0.0712
## 29 wildfire_annualized_frequency                               -0.0935
## 30 wildfire_expected_annual_loss_rate_agriculture              -0.112
## 31 wildfire_historic_loss_ratio_population                     -0.143
## 32 wildfire_expected_annual_loss_rate_building                 -0.197
## 33 wildfire_expected_annual_loss_rate_national_percentile      -0.203
## 34 wildfire_expected_annual_loss_rate_population               -0.241
## 35 wildfire_historic_loss_ratio_buildings                      NA
```




```r
library(formatR)
library(knitr)
library(tidyverse)
library(stargazer)
library(ggcorrplot)
opts_chunk$set(echo = FALSE, eval = TRUE, tidy = TRUE, warning = FALSE, message = FALSE)
Risk_H2AWorkers <- read_csv("~/internship/workspace/wildfire_disaster.csv")

# initial correlation tests
temp_cor_test <- cor.test(Risk_H2AWorkers$mean_temperature, Risk_H2AWorkers$total_workers,
    method = "spearman")
fire_exp_cor_est <- cor.test(Risk_H2AWorkers$wildfire_exposure_total, Risk_H2AWorkers$total_workers,
    method = "spearman")
fire_hazard_cor_tset <- cor.test(Risk_H2AWorkers$wildfire_hazard_type_risk_index_score,
    Risk_H2AWorkers$total_workers, method = "spearman")

# seeing what the data looks like
plot1 <- ggplot(data = Risk_H2AWorkers, aes(y = total_workers, x = wildfire_hazard_type_risk_index_score
    geom_point(aes(color = county_name)) + geom_smooth(method = "lm", se = FALSE,
    col = "black", linewidth = 0.5) + facet_wrap(. ~ state, scales = "free") + theme_minimal() +
    theme(legend.position = "none") + labs(title = "Wildfire Risk Score and Worker Totals",
    x = "Wildfire Hazard Type Risk Index Score", y = "Total H2A Workers")

plot2 <- ggplot(data = Risk_H2AWorkers, aes(y = total_workers, x = mean_temperature)) +
    geom_point(aes(color = county_name)) + geom_smooth(method = "lm", se = FALSE,
    col = "black", linewidth = 0.5) + facet_wrap(. ~ state, scales = "free") + theme_minimal() +
```

```r
    theme(legend.position = "none") + labs(title = "Average Temperature and Worker Totals",
    x = "Mean Temperature", y = "Total H2A Workers")

plot1
plot2

split_datasets <- split(Risk_H2AWorkers, Risk_H2AWorkers$state)

MaxTemp_WorkerTotal_lm <- lm(total_workers ~ maximum_temperature, data = Risk_H2AWorkers)

colorado <- split_datasets[[1]]
montana <- split_datasets[[2]]
north_dakota <- split_datasets[[3]]
south_dakota <- split_datasets[[4]]
utah <- split_datasets[[5]]
wyoming <- split_datasets[[6]]

# maximum temperature correlated with worker totals
colorado_lm <- lm(maximum_temperature ~ total_workers, data = colorado)
wyoming_lm <- lm(maximum_temperature ~ total_workers, data = wyoming)
north_dakota_lm <- lm(maximum_temperature ~ total_workers, data = north_dakota)
south_dakota_lm <- lm(maximum_temperature ~ total_workers, data = south_dakota)
utah_lm <- lm(maximum_temperature ~ total_workers, data = utah)

CO_rsq <- cor(x = colorado$maximum_temperature, y = colorado$total_workers)
MT_rsq <- cor(x = montana$maximum_temperature, y = montana$total_workers)
ND_rsq <- cor(x = north_dakota$maximum_temperature, y = north_dakota$total_workers)
SD_rsq <- cor(x = south_dakota$maximum_temperature, y = south_dakota$total_workers)
WY_rsq <- cor(x = wyoming$maximum_temperature, y = wyoming$total_workers)
UT_rsq <- cor(x = utah$maximum_temperature, y = utah$total_workers)

r_squared <- data.frame(State = c("Colorado", "Wyoming", "North Dakota", "South Dakota",
    "Utah"), Model_R_sqr = c(summary(colorado_lm)$r.squared, summary(wyoming_lm)$r.squared,
    summary(north_dakota_lm)$r.squared, summary(south_dakota_lm)$r.squared, summary(utah_lm)$r.squared)
    Pearson_Corr = c(CO_rsq, WY_rsq, ND_rsq, SD_rsq, UT_rsq))

kable(r_squared, col.names = c("State", "Multiple R-Sqr", "Pearson Corr. Coef."),
    align = "lcc", label = "Correlation Between Heat and Worker Density", digits = 3)

FireHzrdIdxScore_WorkerTotal_lm <- lm(total_workers ~ wildfire_hazard_type_risk_index_score,
    data = Risk_H2AWorkers)
FireExpTotal_WorkerTotal_lm <- lm(total_workers ~ wildfire_exposure_total, data = Risk_H2AWorkers)

cat("Linear Model of Maximum Temperature and Worker Totals: \n")
stargazer(FireExpTotal_WorkerTotal_lm, type = "text", title = "Summary Statistics",
    out = "summary_statistics.txt")
cat("\nLinear Model of Wild Fire Hazard Index Score and Worker Totals: \n")
stargazer(FireHzrdIdxScore_WorkerTotal_lm, type = "text", title = "Summary Statistics of Coefficients",
    out = "summary_coefficients.txt")
cat("\nLinear Model of Fire Exposure Total and Worker Totals: \n")
stargazer(FireExpTotal_WorkerTotal_lm, type = "text", title = "Summary Statistics of Coefficients",
    out = "summary_coefficients.txt")
```

```r
# calculating all correlations with total workers
correlation_vector <- c()
Risk_H2AWorkers_num <- Risk_H2AWorkers[, sapply(Risk_H2AWorkers, is.numeric)]
N <- colnames(Risk_H2AWorkers_num)
for (i in N) {
    correlation_value <- cor(x = Risk_H2AWorkers_num[[i]], y = Risk_H2AWorkers_num$total_workers,
        method = "spearman")
    correlation_vector <- c(correlation_vector, correlation_value)
}
correlation_df <- data.frame(column = N, correlation = correlation_vector) |>
    arrange(desc(correlation))
correlation_df <- correlation_df[-1, ]

# Correlogram
temp <- Risk_H2AWorkers[sapply(Risk_H2AWorkers, is.numeric)]
corr <- round(cor(temp, method = "spearman"), 2)
matrix <- cor_pmat(temp)
corrplot <- ggcorrplot(corr, p.mat = matrix, type = "full", lab = TRUE, lab_size = 1.6,
    insig = "blank", title = "Correlation Plot") + theme(axis.text.x = element_text(size = 4.1,
    angle = 90, hjust = 1), axis.text.y = element_text(size = 4.1))

cat("\n\n Correlation of all Variables with Total Workers \n")
print(tibble(correlation_df), n = nrow(correlation_df))
cat("\n\n")
corrplot
```