Untitled

```r
rdu <- read_csv('weather_station_data_20220915/rdu_station_data.csv')
# eco <- read_csv('weather_station_data_20220915/econet_weather_station_data.csv')
# roxboro
# tdf <- read_csv('weather_station_data_20220915/tdf_station_data.csv')
```

```r
# get year, month, day from date

# split into separate columns
date <- unique(rdu$date)

get_date_component <- function(x, n) {
  res <- str_split(x, '/')
  return(res[[1]][n])
}

month <- sapply(date, get_date_component, n=1)
day <- sapply(date, get_date_component, n=2)
year <- sapply(date, get_date_component, n=3)

# make df
dates_df <- data.frame(date, month, day,year, row.names=NULL)

rdu %>%
  left_join(dates_df, by='date') -> rdut

# make factor
rdut$month <- as.factor(rdut$month)
rdut$day <- as.factor(rdut$day)
rdut$year <- as.factor(rdut$year)

# MAKES RDUT DF
```

```r
# add extreme heat condition var
rdut %>%
  mutate(eh = case_when(
    (hi >= 124) ~ 'Extreme Danger',
    (hi >= 103) & (hi < 124) ~ 'Danger',
    (hi >= 90) & (hi < 103) ~ 'Extreme Caution',
    (hi >= 80) & (hi < 90) ~ 'Caution',
    TRUE ~ "None"
  )) -> rdut
```
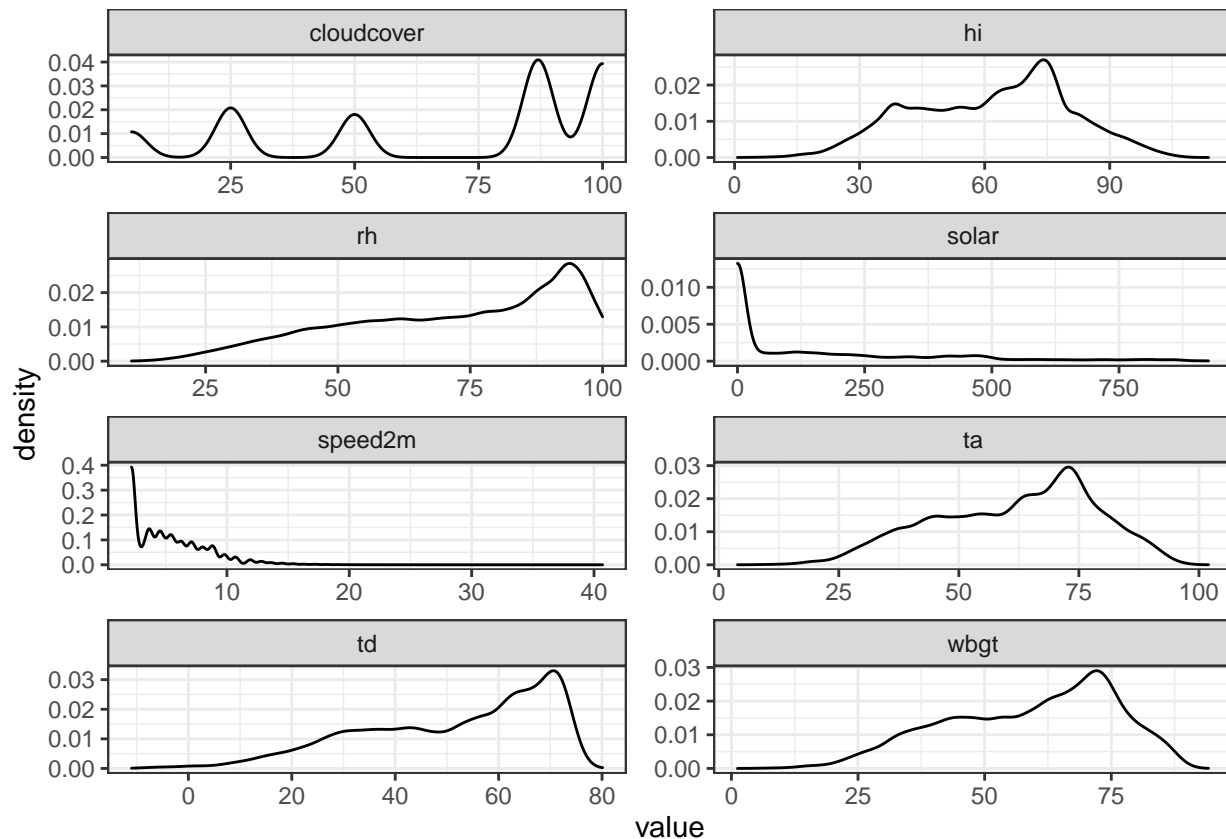
summary stats

density by factor

```
# there are no extreme danger points
rdut %>%
  count(eh)
```

```
## # A tibble: 4 x 2
##   eh                 n
##   <chr>          <int>
## 1 Caution         8979
## 2 Danger           303
## 3 Extreme Caution 4490
## 4 None           71573
```

```
# histogram by each variable overall
rdut %>%
  select(wbgt:hi, year, month, day)%>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value)) +
  geom_density() +
  facet_wrap(~type, nrow=4, scales='free')+
  theme_bw()
```
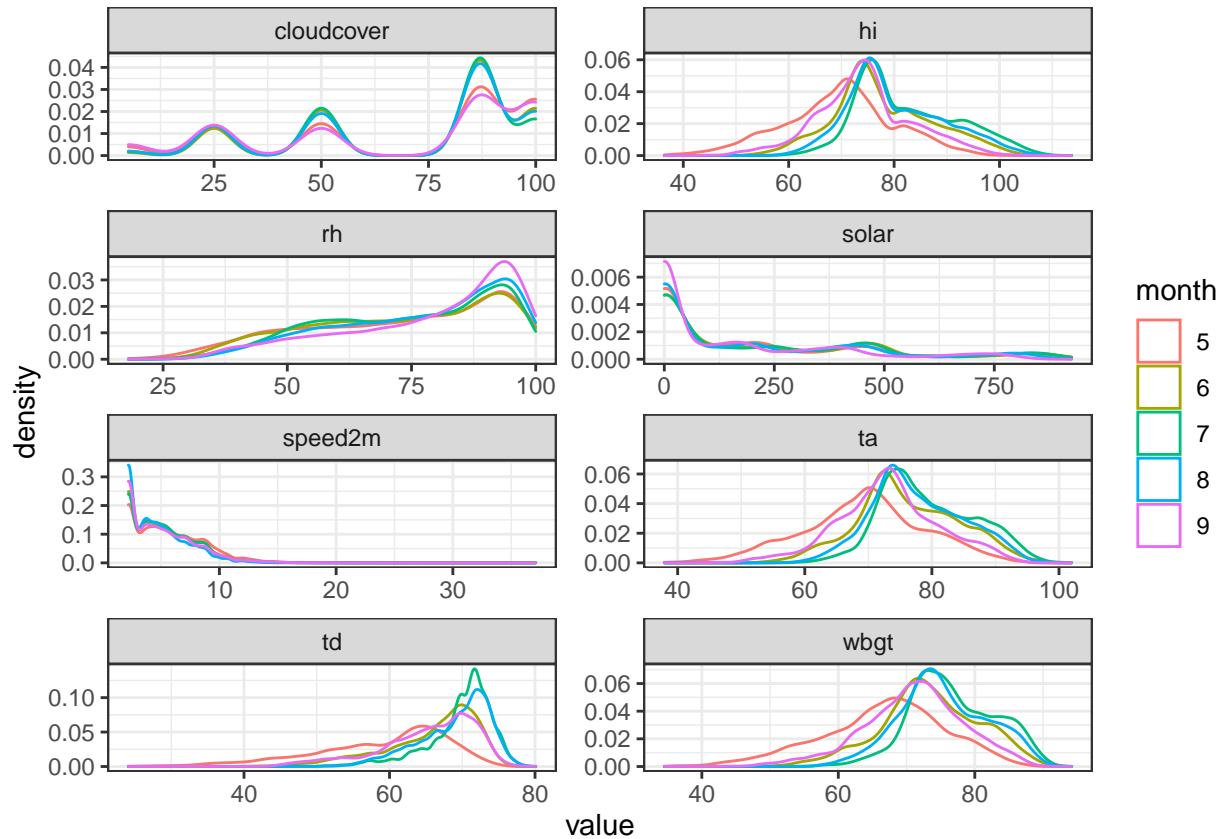

density by month for summer

```
rdut %>%
  select(wbgt:hi, year, month, day) %>%
  filter(month %in% c(5,6,7,8,9)) %>%
```

```
pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
ggplot(aes(x = value, color = month)) +
geom_density() +
facet_wrap(~type, nrow=4, scales='free') +
theme_bw()
```
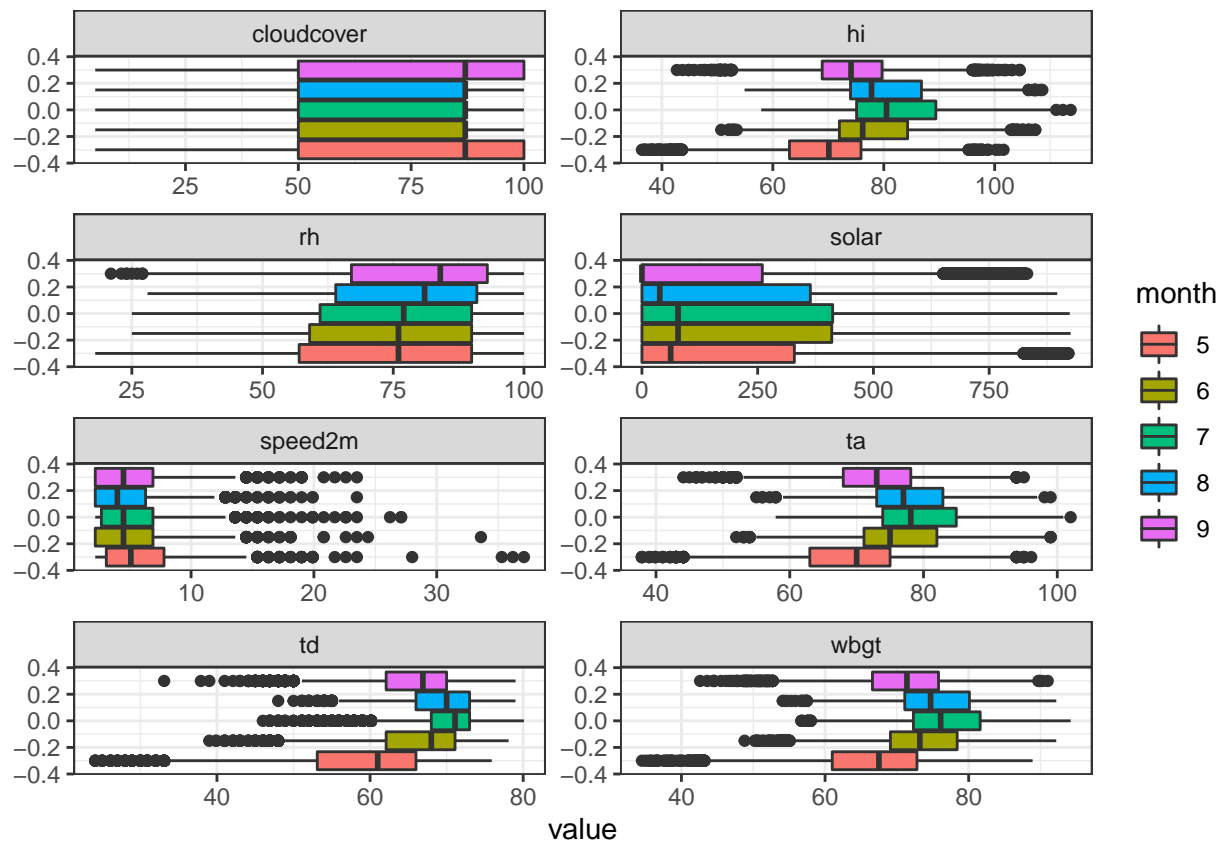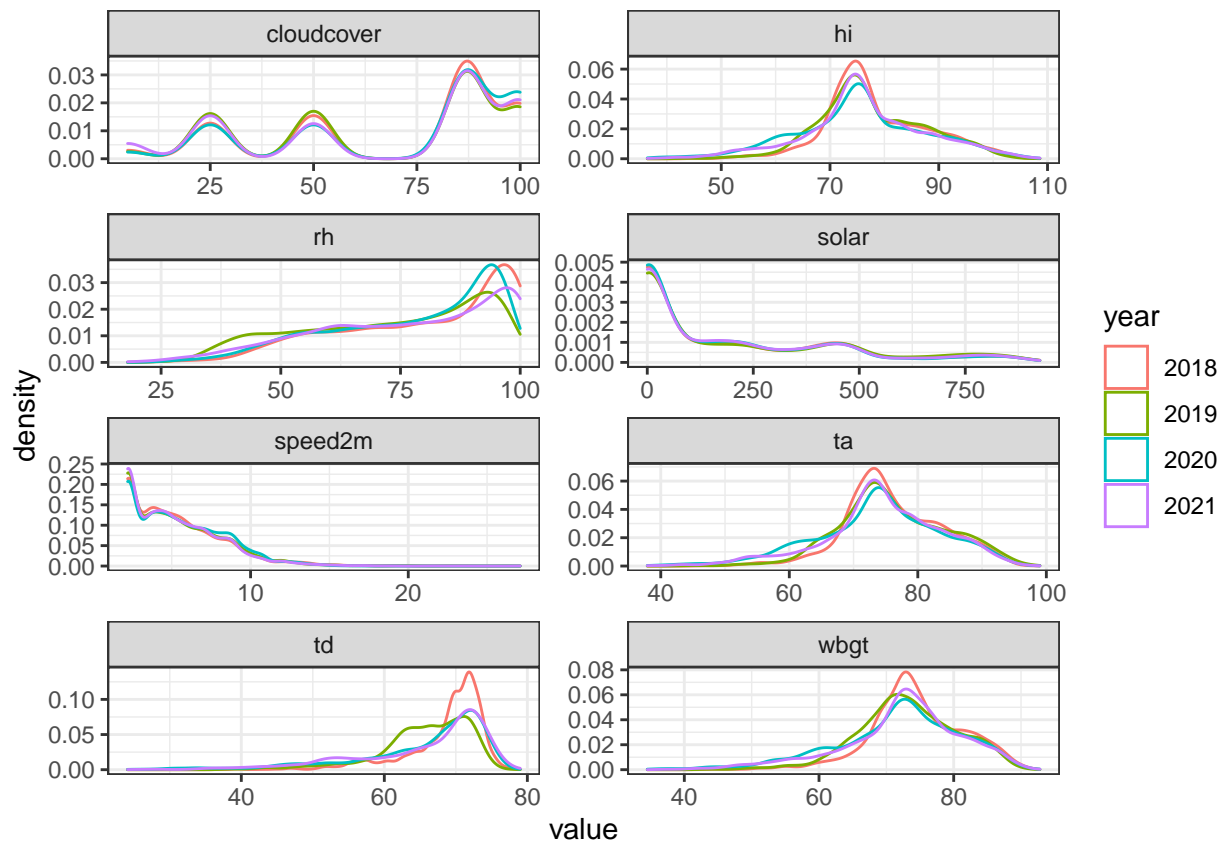


```
rdut %>%
  select(wbgt:hi, year, month, day) %>%
  filter(month %in% c(5,6,7,8,9)) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, fill = month)) +
  geom_boxplot() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw()
```

density by year for summer

```
rdut %>%
  select(wbgt:hi, year, month, day) %>%
  filter(month %in% c(5,6,7,8,9),
         year %in% c(2018, 2019, 2020,2021)) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, color = year)) +
  geom_density() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw()
```
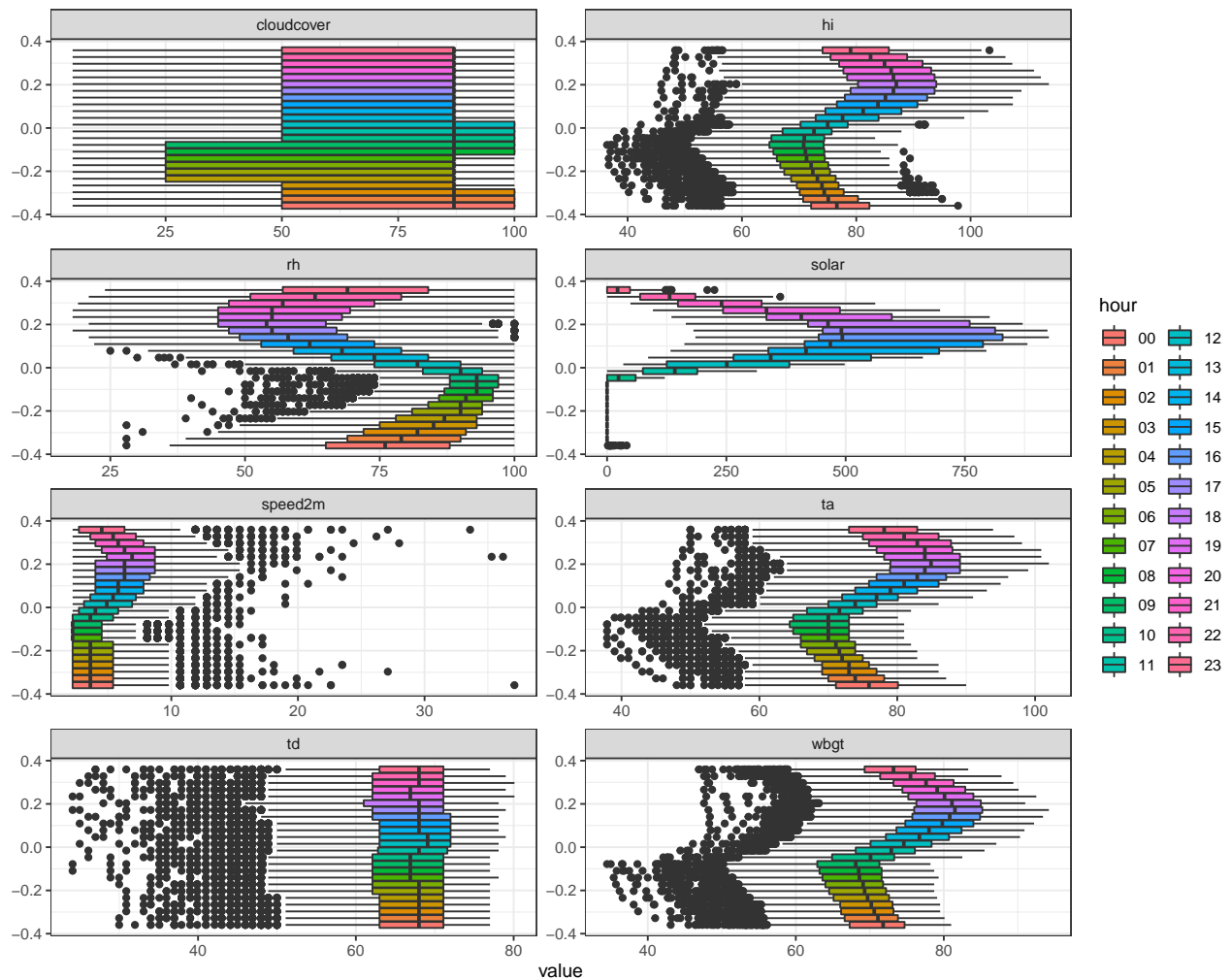
```
rdut %>%
  select(wbgt:hi, year, month, day) %>%
  filter(month %in% c(5,6,7,8,9),
         year %in% c(2018, 2019, 2020,2021)) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, fill = year)) +
  geom_boxplot() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw()
```
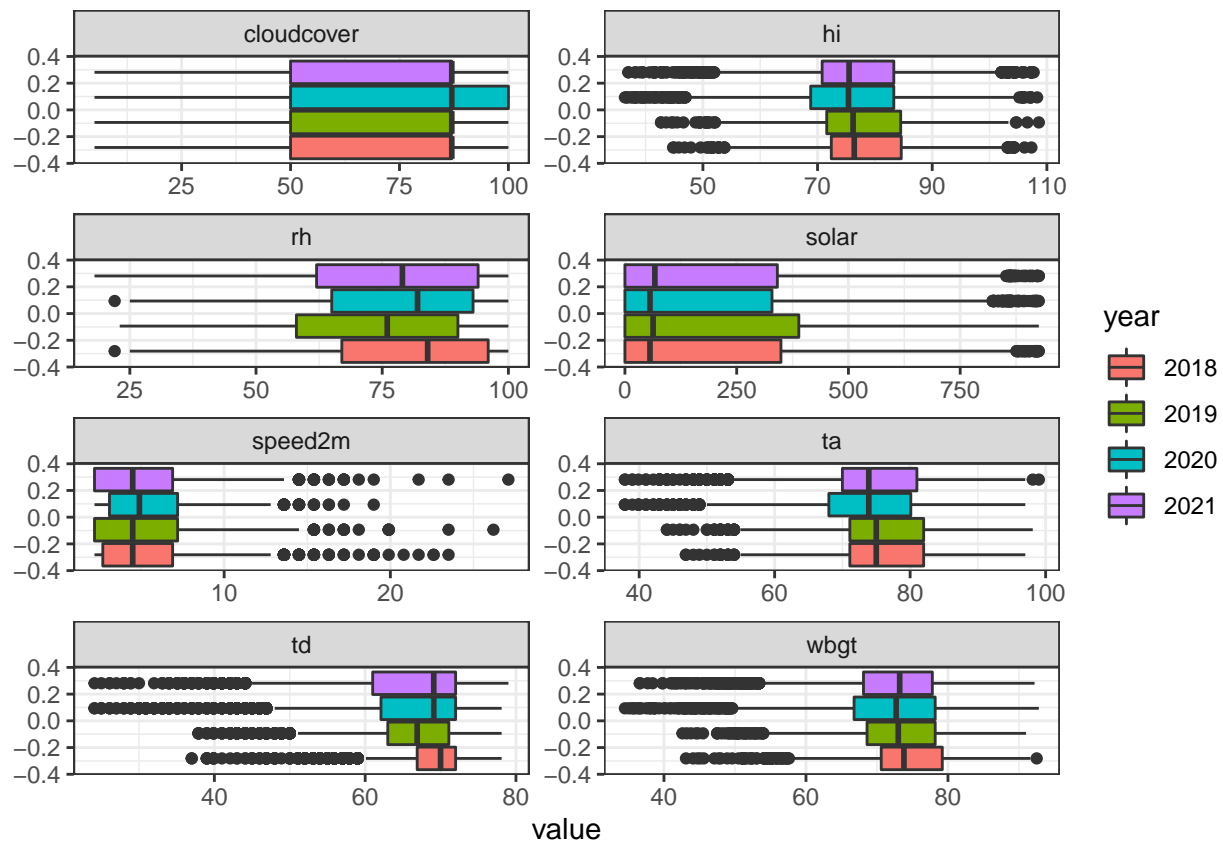
distribution of factor by time of day for summer months

```r
rdut$time2 <- as.character(rdut$time)
rdut %>%
  mutate(hour = as.factor(str_sub(time2, 1,2))) %>%
  select(wbgt:hi, year, month, day, hour) %>%
  filter(month %in% c(5,6,7,8,9)) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, fill = hour)) +
  geom_boxplot() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw()
```

```r
# make heat warning boolean variable
rdut %>%
  mutate(is_eh = ifelse(eh == 'None', FALSE, TRUE)) -> rdut
```

compare heat warning vs not for each factor

```r
rdut %>%
  filter(month %in% c(5,6,7,8,9)) %>%
  select(wbgt:hi, is_eh) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, color = is_eh)) +
  geom_density() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw()
```
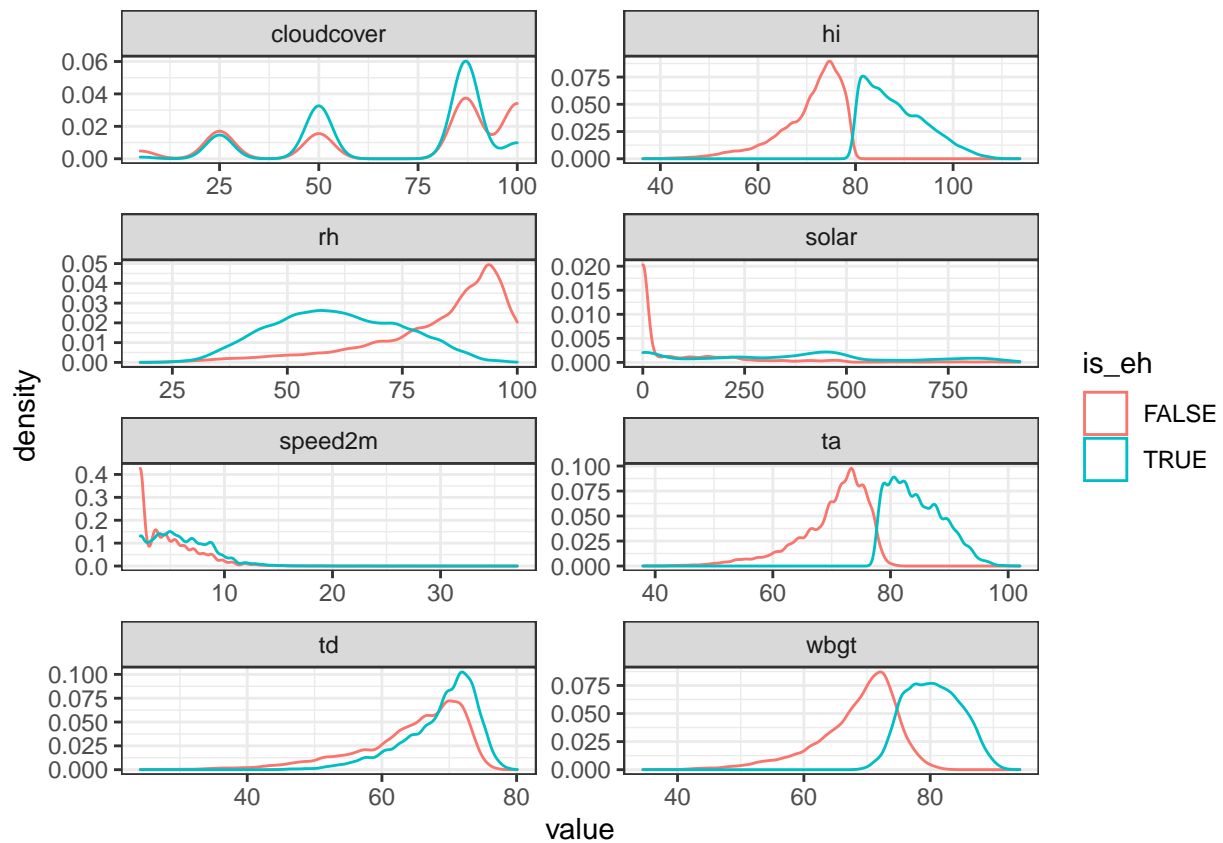
```
rdut %>%
  filter(month %in% c(5,6,7,8,9)) %>%
  select(wbgt:hi, is_eh) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = value, color = is_eh)) +
  geom_boxplot() +
  facet_wrap(~type, nrow=4, scales='free') +
  theme_bw() +
  labs(color = 'Heat Index Warning')
```
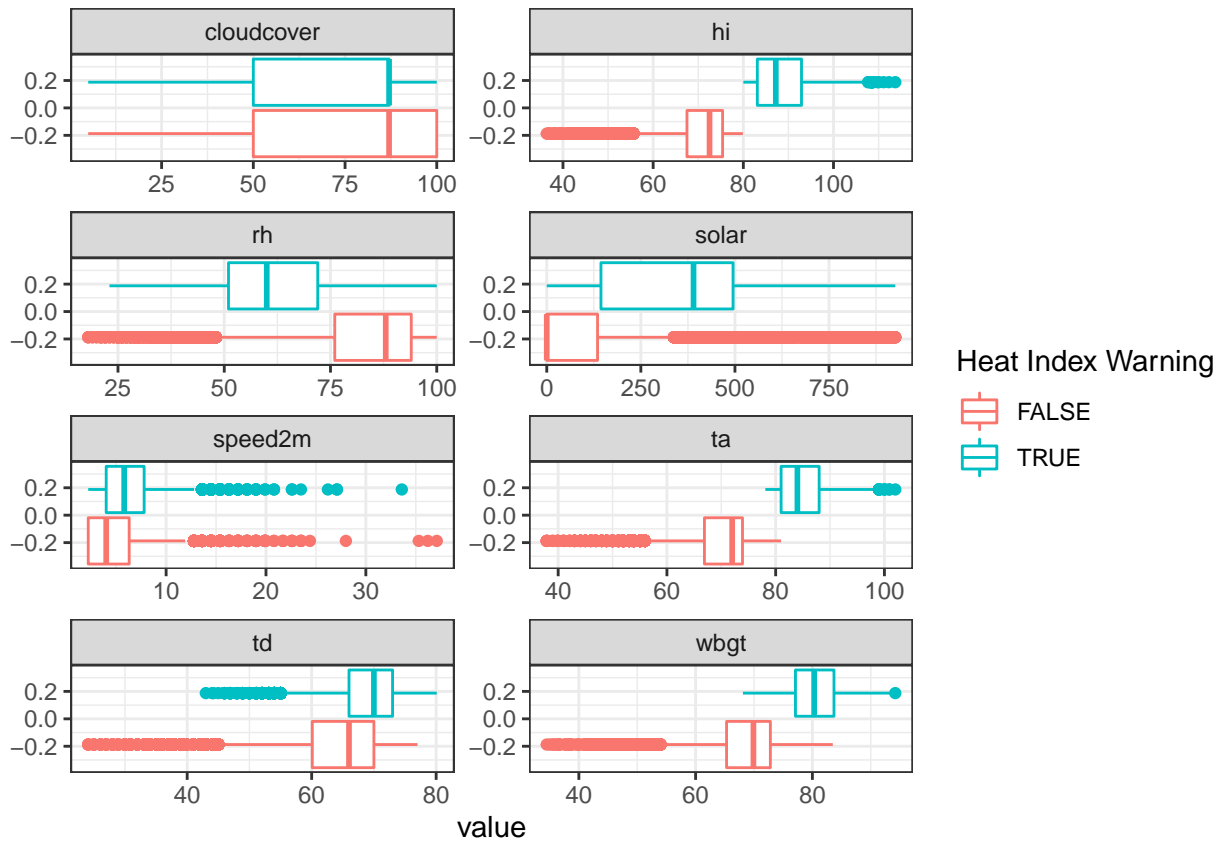
```
rdut$day <- factor(rdut$day, levels=c(1:31))


rdut %>%
  filter(month %in% c(5,6,7,8,9)) %>%
  mutate(hour = as.factor(str_sub(time2, 1,2))) %>%
  mutate(tod = case_when(
    hour %in% c('22', '23', '00', '01', '02', '03', '04', '05') ~ 'Evening',
    hour %in% c('06', '07', '08', '09', '10', '11') ~ 'Morning',
    hour %in% c('12', '13','14','15','16','17') ~ 'Afternoon',
    hour %in% c('18','19','20','21') ~ 'Early Evening'
  )) -> rdut_tod

rdut_tod$tod <- factor(rdut_tod$tod, levels = c('Morning', 'Afternoon', 'Early Evening', 'Evening'))

rdut_tod %>%
  filter(is_eh == TRUE) %>%
  select(wbgt:hi, month:year, is_eh, tod) %>%
  group_by( month, day, tod) %>%
  summarise_if(is.numeric, mean) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = day, y = value, color = tod)) +
  geom_line(aes(group = 1)) +
  facet_grid(type~month, scales = 'free') +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  labs(color='Time of Day')
```
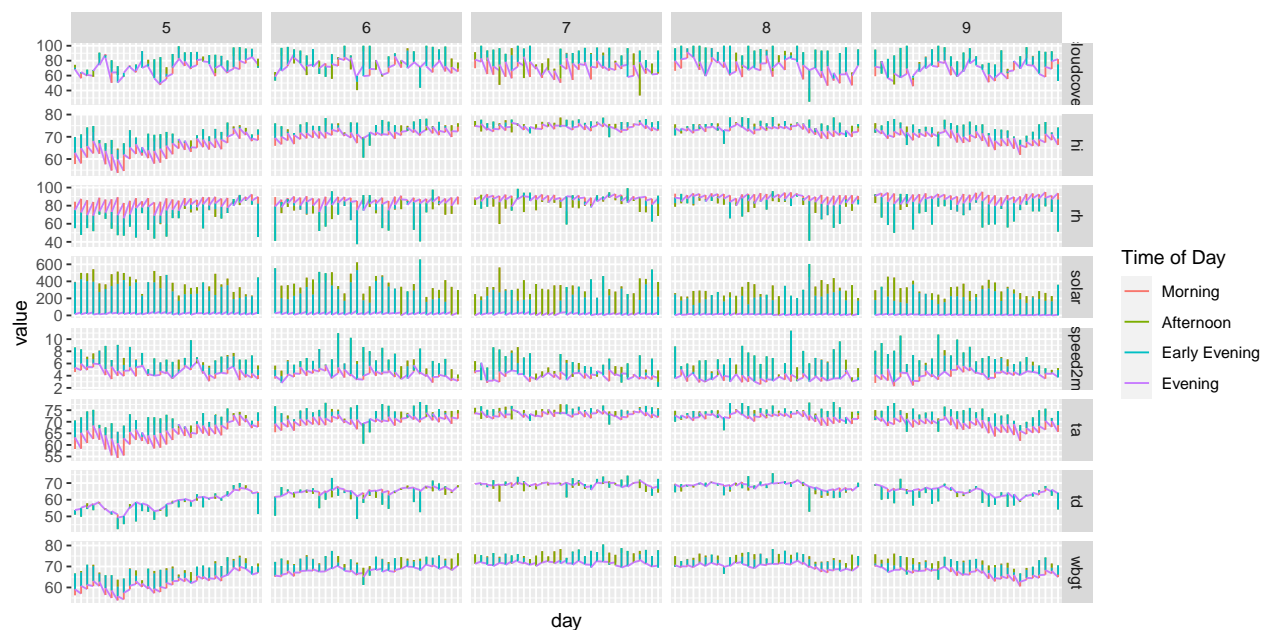
9
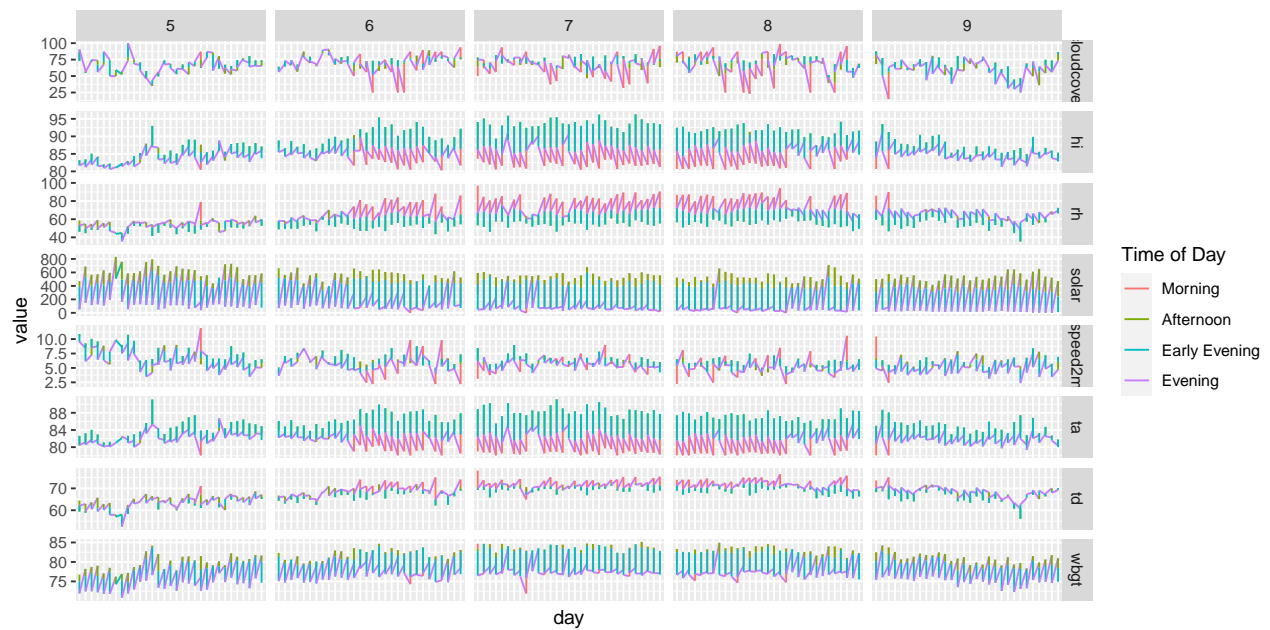
```
rdut_tod %>%
  filter(is_eh == FALSE) %>%
  select(wbgt:hi, month:year, is_eh, tod) %>%
  group_by( month, day, tod) %>%
  summarise_if(is.numeric, mean) %>%
  pivot_longer(cols=wbgt:hi, names_to='type', values_to = 'value') %>%
  ggplot(aes(x = day, y = value, color = tod)) +
  geom_line(aes(group = 1)) +
  facet_grid(type~month, scales = 'free') +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  labs(color='Time of Day')
```


look at lag