ESPM 205 - Homework assignment 2

Jennifer Natali, with help from chatgpt for learning R and dplyr

2024 October 16 [extremely late]

Instructions

Work on the exercise below, due October 1st (before midnight) on bCourses. Please write **short answers (<100 words)** in "your text here", as well as the R code you used to get there (in "your code here").

Exercise

1) Please upload a sample of the data (.csv file) you plan to analyze for your final project. You can (but do not need to) upload the whole dataset. However, the closer the resemblance of this data set to the one you will end up analyzing, the better. E.g., if your question is at the community level, then include several species; if you would like to compare a particular physical variable across different sites, then include several sites. The goal is for you to start getting familiar with your data and its level of complexity. In the code below, import your data set in R, and examine the following properties: (1) length and frequency of the time series (whether it is one, or multiple time series); (2) completeness of each time series; (3) basic descriptive statistics for each time series (at least mean, CV, ACF for each variable; plus anything else you would like to add). [6 points total]

```
# Load libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(astsa)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
# Setup directories and filepaths
home_dir='/Volumes/SANDISK_SSD_G40/GoogleDrive/GitHub/'
repository_dir = paste(home_dir,'sagehen_meadows/', sep='')
groundwater_filepath = paste(repository_dir,'data/field_observations/groundwater/biweekly_manual/groundw

# Load groundwater data
```

```r
groundwater <- read.csv(groundwater_filepath)

# summarize the full times series
summary(groundwater)
```

```
##    well_id           timestamp         ground_to_water_cm
##  Length:1116        Length:1116        Min.   :-35.41
##  Class :character   Class :character   1st Qu.: 18.30
##  Mode  :character   Mode  :character   Median : 41.86
##                                        Mean   : 43.85
##                                        3rd Qu.: 63.59
##                                        Max.   :194.67
##                                        NA's   :151
```

```r
# length of full time series by date
groundwater <- groundwater %>%
  mutate(timestamp = as.POSIXct(timestamp, format = "%m/%d/%Y %H:%M"))

groundwater <- groundwater %>% mutate(date = as.Date(timestamp))
groundwater %>% summarize(
  start_date = min(date, na.rm=TRUE),
  stop_date = max(date, na.rm=TRUE),
  timespan = difftime(stop_date, start_date, units="days")
)
```

```
##   start_date  stop_date  timespan
## 1 0018-05-31 0021-11-14 1263 days
```

```r
# number of unique dates in the full time series
groundwater %>%
  summarize(unique_dates_count = n_distinct(date))
```

```
##   unique_dates_count
## 1                 59
```

```r
# length of full time series by week of the year
# -- NOTE: some well readings are on consecutive days, may want to group them
#          without altering the date; but doing this by week won't work!!
#          (e.g. 11/18/19 and 11/19/19 are different isoweeks)

# groundwater <- groundwater %>% mutate(week = isoweek(date))
# groundwater_weekly_summary <- groundwater %>%
#   group_by(week) %>%
#   summarize(
#     start_week = min(week),
#     end_week = max(week),
#     n_week = n()  # Number of entries in each week
#   )
# groundwater_weekly_summary

# completeness of full time series in terms of entries with NA
# ---NOTE: expect 153 NA entries due to NO WATER readings
# ---TODO: create a ground_to_water_greater_than column for NO WATER readings
#          to capture the depth of the well; this is more info than nothing.
groundwater %>%
  summarize(
```

```
    na_sum = sum(is.na(ground_to_water_cm)),
    na_percent = 100 * sum(is.na(ground_to_water_cm)) / n())
```

```
##   na_sum na_percent
## 1    151   13.53047
```

```
# basic descriptive statistics across all groundwater readings: mean, CV, ACF
groundwater %>%
  summarize(
    mean_value = mean(ground_to_water_cm, na.rm = TRUE),
    sd_value = sd(ground_to_water_cm, na.rm = TRUE),
    var_value = var(ground_to_water_cm, na.rm = TRUE),
    cv_value = 100 * sd_value / mean_value
  )
```

```
##   mean_value sd_value var_value cv_value
## 1   43.84532 34.49195  1189.695 78.66735
```

```
groundwater_by_day <- groundwater %>%
  group_by(date) %>%
  summarise(
    mean_value = mean(ground_to_water_cm, na.rm = TRUE),
    sd_value = sd(ground_to_water_cm, na.rm = TRUE),
    var_value = var(ground_to_water_cm, na.rm = TRUE)
  )

# basic descriptive statistics by groupings: mean, CV, ACF for each variable
groundwater_by_well <- groundwater %>%
  group_by(well_id) %>%
  summarise(
    mean_value = mean(ground_to_water_cm, na.rm = TRUE),
    sd_value = sd(ground_to_water_cm, na.rm = TRUE),
    var_value = var(ground_to_water_cm, na.rm = TRUE)
  )
groundwater_by_well
```

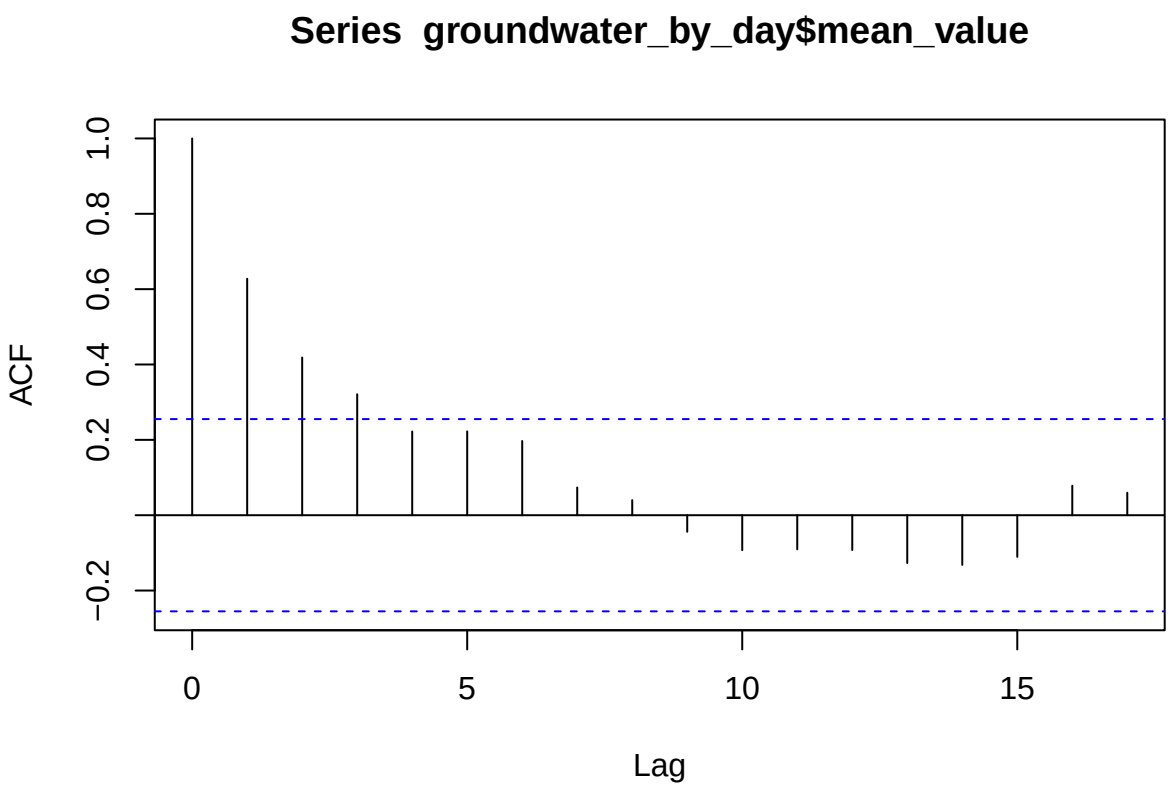
```
## # A tibble: 54 x 4
##    well_id  mean_value sd_value var_value
##    <chr>         <dbl>    <dbl>     <dbl>
##  1 EEF-1          28.8    31.7     1003.
##  2 EER-1          34.1    10.8      117.
##  3 EET-1          59.9    32.8     1076.
##  4 EET-2         101.     41.2     1699.
##  5 EET-XB4S       31.6    20.0      400.
##  6 EFF-XA1N       40.1    11.3      127.
##  7 EFF-XA2N       27.0    11.6      134.
##  8 EFF-XB7S       30.5    36.8     1357.
##  9 EFR-XB1S       45.2     4.96      24.6
## 10 EFR-XB2N       28.7    11.3      127.
## # i 44 more rows
```

```
groundwater_by_well_year <- groundwater %>%
  mutate(year = year(timestamp)) %>%
  group_by(well_id, year) %>%
  summarise(
    mean_value = mean(ground_to_water_cm, na.rm = TRUE),
```

```
    sd_value = sd(ground_to_water_cm, na.rm = TRUE),
    var_value = var(ground_to_water_cm, na.rm = TRUE),
    .groups = "keep"
  )
groundwater_by_well_year
```

```
## # A tibble: 141 x 5
## # Groups:   well_id, year [141]
##    well_id  year mean_value sd_value var_value
##    <chr>   <dbl>      <dbl>    <dbl>     <dbl>
##  1 EEF-1      18       6.21     7.70      59.3
##  2 EEF-1      19      19.2     21.6      465.
##  3 EEF-1      21      61.0     29.8      888.
##  4 EER-1      18      34.5      8.50      72.3
##  5 EER-1      19      27.6     12.4      155.
##  6 EER-1      21      39.1      8.76      76.7
##  7 EET-1      18      65.4     31.7     1007.
##  8 EET-1      19      57.0     41.7     1736.
##  9 EET-1      21      56.4     31.1      968.
## 10 EET-2      19     107.      20.4      418.
## # i 131 more rows
```

```
# Trying to get acf, but not sure if this summarized data means anything
acf(groundwater_by_day$mean_value)
```



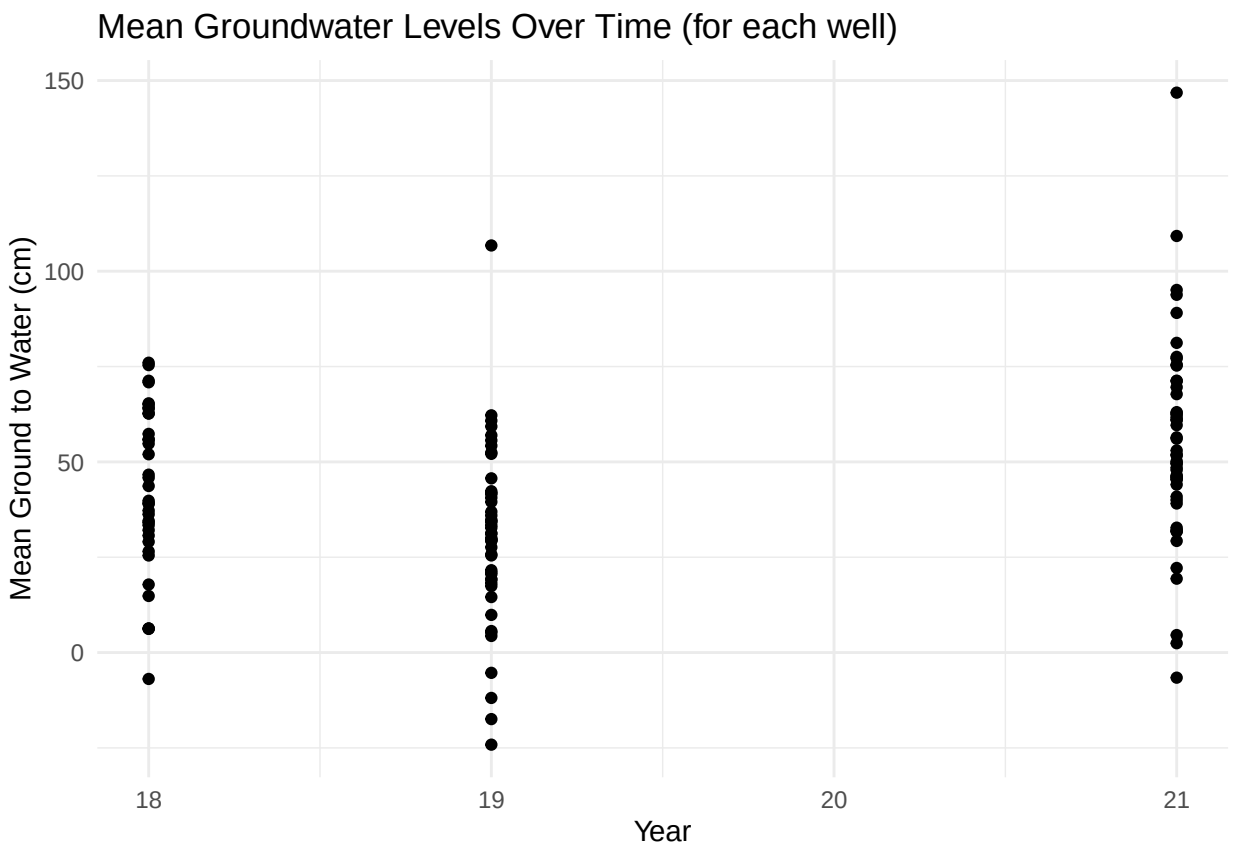

```
# TODO Next step summary statistics
# ------wells in each meadow group: kiln, east, low
# ------wells from each plant functional type: sedge, willow, mixed herbaceous, pine
# ------wells from each hydrogeomorphic zone: riparian, terrace, fan
```

```
# TODO: Next time series to consider
# ------discharge (at one point)
# ------daily precipitation (at one point)
# ------sunlight, aka PAR (at one point)
# ------max, mean daily temperature (at each meadow)
```

2) Plot the data [2 points].

```
# Load libraries
library(ggplot2)
library(tidyr)

# Plot (1): the mean annual groundwater level for all wells for each year
ggplot(groundwater_by_well_year, aes(x = year, y = mean_value, group = well_id)) +
  geom_point() +  # Optional: add points at each data point
  labs(title = "Mean Groundwater Levels Over Time (for each well)",
       x = "Year", y = "Mean Ground to Water (cm)") +
  theme_minimal()
```



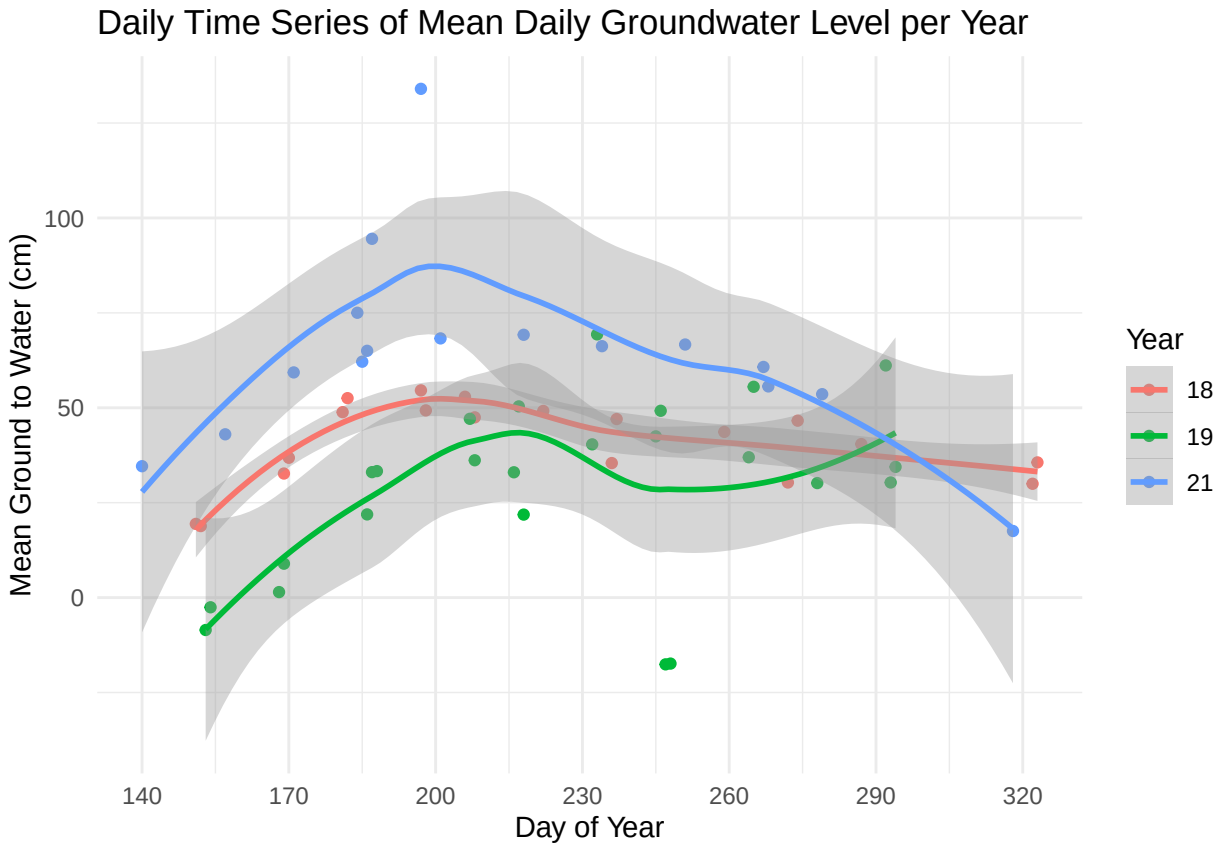

```
# Plot (2): year-over-year daily time series of mean groundwater level

# ---Setup dataframe with new columns for year and day_of_year
groundwater_by_day <- groundwater_by_day %>%
  mutate(
    year = year(date),                # Extract the year from Date
    day_of_year = yday(date)          # Extract the day of year (1-365/366)
  )
```

```r
# ---Add NA values for days with no measurement (or mean_value)
complete_groundwater_by_day <- groundwater_by_day %>%
  group_by(year) %>%
  complete(day_of_year =
           min(groundwater_by_day$day_of_year):max(groundwater_by_day$day_of_year),
           fill = list(mean_value = NA))  # Fill missing days with NA

# ---Plot it!
ggplot(complete_groundwater_by_day, aes(x = day_of_year, y = mean_value, color = factor(year), group = y
  geom_point() +
  geom_smooth() +
  theme_bw() +
  labs(title = "Daily Time Series of Mean Daily Groundwater Level per Year",
       x = "Day of Year",
       y = "Mean Ground to Water (cm)",
       color = "Year") +
  scale_x_continuous(breaks = seq(min(groundwater_by_day$day_of_year),
                                  max(groundwater_by_day$day_of_year), by = 30)) +  # Customize x-axis
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



3) What is your main research question? Do you have any working hypothesis? [2 points].

Q1: How does meadow groundwater vary by season and climate as influenced by elevation, hydrogeomorphic zones, and evapotranspiration rates of plant functional types? I expect evapotranspiration to drive daily and seasonal groundwater levels with sensitivity to meteorology and day length. Q2: What controls plant

functional type phenology? I hypothesize that peak productivity and senescence will correlate to groundwater levels as governed by meteorology but moderated by hydrogeomorphic zones and elevation. Q3: Does topography or subsurface character influence groundwater reliability? I expect that groundwater reliability will correlate to topographic convergence or subsurface boundaries (i.e. differing conductivity).

Any notes (optional)

Thank you for patience and support. So excited to be digging into this data with renewed energy and focused guidance! Been dreaming of this day.