Analyzing Soil Moisture & Precipitation Trends in Coweeta Basin LTER Site

Web address for GitHub repository

Kelly Davidson, Megan McClaugherty, & Isabel Zungailia

Contents

1 Rationale and Research Questions **5**

2 Dataset Information **8**

3 Exploratory Analysis **11**

4 Analysis **20**

4.1 Question 1: Has soil moisture in the upper 30 centimeters of soil increased or decreased in Coweeta Basin from 2000 to 2013 at each of the four research sites? 20

4.2 Question 2: Which sites experienced the greatest change in soil moisture over time? 20

4.3 Question 3: What is the impact of elevation on soil moisture? 20

4.4 Question 4: Is there a relationship between soil moisture and precipitation that can be observed at these four research sites? 23

5 Summary and Conclusions **25**

6 References **26**

List of Tables

List of Figures

1 Soil moisture research sites and rain gauge locations within Coweeta Basin, North Carolina. 6
2 Processed soil moisture data. 9
3 Processed rain gauge/precipitation data. 10

1 Rationale and Research Questions

The Coweeta Basin Long Term Ecological Research (LTER) site is deemed one of the oldest continuous environmental studies in North America. It is managed by the USDA Forest Service in collaboration with the University of Georgia. This site is located in the southern portion of the Appalachian Mountains in North Carolina and is primarily composed of deciduous forest. We decided to use this site for our research project because it provides a unique setting to explore the relationships between environmental variables such as soil moisture, precipitation, and elevation over time. The goal of this analysis was to use skills that we have learned throughout the semester to generate and answer a specific research question. This investigation was intended to analyze temporal trends in soil moisture throughout the Coweeta Basin, and we used soil moisture data from four sites (at four different elevations) to explore our research questions. We also investigated the relationship between precipitation and soil moisture over time, assuming that there is a positive relationship between these two variables. The analysis is based on data collected between the years 2000 and 2013. The four sites where soil moisture data was collected are located at four different elevations, allowing us to infer the relationship between soil moisture and elevation. The research sites are located at the following elevations:

- Site 1: 817.17 meters
- Site 2: 1380.44 meters
- Site 3: 1214.93 meters
- Site 4: 735.47 meters

The Coweeta LTER laboratory collects precipitation data from a total of eight rain gauges throughout the Coweeta Basin, so we used the geographic coordinates of the rain gauges and the four research sites to determine which rain gauge was located closest to each site. The precipitation data from Rain Gauge 6 (RG06) was used for Sites 1 and 4, data from Rain Gauge 31 (RG31) was used for Site 2, and data from Rain Gauge 5 (RG05) was used for Site 3. All of the sites are within 0.8 km of their paired rain gauge.

Our main research question: *Has soil moisture in the upper 30 cm (of soil) increased or decreased in Coweeta Basin from 2000 to 2013 at each of the four research sites?*

Other questions that were explored include:

- *Which sites experienced the greatest change in soil moisture over time?*
- *What is the impact of elevation on soil moisture?*
- *Is there a relationship between soil moisture and precipitation that can be observed at these four study sites?*

Null Hypothesis A: There was no change in soil moisture in the upper 30 cm from 2000 to 2013 at each of the four research sites. **Alternative Hypothesis A**: There was a change in soil moisture in the upper 30 cm from 2000 to 2013 at each of the four research sites.

Null Hypothesis B: There is no relationship between soil moisture and precipitation at each of the four research sites. **Alternative Hypothesis B**: There is a relationship between soil moisture and precipitation at each of the four research sites.

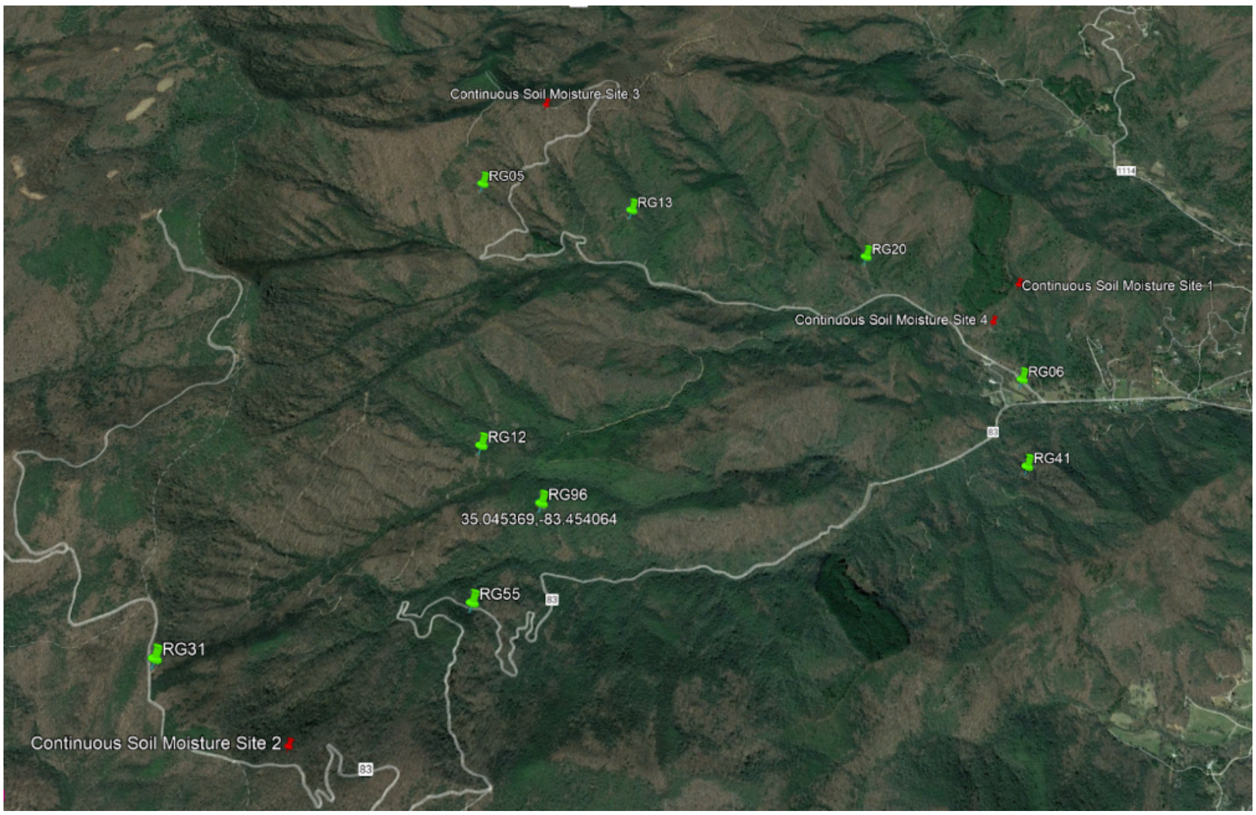


Figure 1: Soil moisture research sites and rain gauge locations within Coweeta Basin, North Carolina.

```r
library(knitr)
```

2 Dataset Information

Data Description

Soil moisture datasets were collected from the EDI Data Portal for Coweeta LTER here. Coweeta LTER (248) was selected as the LTER Site and "Continuously measured forest soil moisture at four sites in the Coweeta Basin" was identified as the dataset of interest. The data was accessed on 11/21/22. The following datasets were downloaded:

- 1040_2000_2014
- 1040.kml (a Google Earth file showing the soil moisture site locations)

Precipitation data was collected from the US Forest Service Southern Research Station here. Daily precipitation data from recording rain gages (RRG) at Coweeta Hydrologic Lab, North Carolina was selected under Coweeta Datasets on the USFS Research Data Archive here. The data was accessed on 11/30/22.

We generated four new datasets combining the monthly average precipitation with monthly average 30 cm soil moisture at the 4 sites monitored at the Coweeta LTER.

Data Wrangling

The initial datasets required a significant amount of wrangling before they could be used for analysis. The raw soil moisture data had 21 variables/columns, so the first step was to select the columns of interest (site, Year, YearDay, smois30). A new 'Date' column was created (format = "%Y-%j") and the data was filtered to exclude the smois30 values less than 0 and greater than 1 since soil moisture was expressed as percent water content. All NAs were omitted from the data and it was then split into 4 separate dataframes, one for each research site. The smois30 (cm) values were averaged per month for each site prior to analysis.

The raw precipitation datasets for the three rain gauges had 5 variables (YEAR, MONTH, DAY, RRG*gauge ID number*). A new 'Date' column was created (format = "%Y-%m-%d"), and the data was wrangled to only include the years 2000 to 2013 that we had soil moisture data for. All NAs were omitted from the data. The final step was to average the rain gauge data by month in order to mirror average monthly soil moisture.

Data Structure

Processed Soil Moisture Data |Variable | Description | Units | Class | Stats | |——————|————|—— |————|————————————-| |Year | Calendar year | |Integer | Minimum = 2000, Maximum = 2013| |——————|——————|——-|————|——————————————-| |Month |Calendar Month | |Integer |Minimum = 01 (January), | | | | | |Maximum = 07 (July) | |——————|——————|——-|————|——————————————-| AverageMonthlySmois30| Average monthly 30 cm soil moisture as percent water content | Unitless (measured as a percent) | Numeric | Site 1: minimum = 0.1521, mean = 0.2804, maximum = 0.3556, Site 2: minimum = 0.1046, mean = 0.2543, maximum = 0.3421, Site 3: minimum = 0.114, mean = 0.2729, maximum = 0.5753, Site 4: minimum = 0.1383, mean = 0.3156, maximum = 0.4922 | |YearMonth| Date in format YYYY-MM-DD | | Character | Minimum = 2000-01-21, Maximum = 2013-07-19|

Variable	**Description**	**Units**	**Class**	**Stats**
Year	Calendar year		Integer	Minimum = 2000 Maximum = 2013
Month	Calendar month		Integer	Minimum = 01 (January) Maximum = 07 (July)
AverageMonthly Smois30	Average monthly 30 cm soil moisture as percent water content	Unitless (measured as a percent)	Numeric	**Site 1**: Minimum = 0.1521 Mean = 0.2804 Maximum = 0.3556
				Site 2: Minimum = 0.1046 Mean = 0.2543 Maximum = 0.3421
				Site 3: Minimum = 0.114 Mean = 0.2729 Maximum = 0.5753
				Site 4: Minimum = 0.1383 Mean = 0.3156 Maximum = 0.4922
YearMonth	Date in format YYYY-MM-DD		Character	Minimum = 2000-01-21 Maximum = 2013-07-19

Figure 2: Processed soil moisture data.

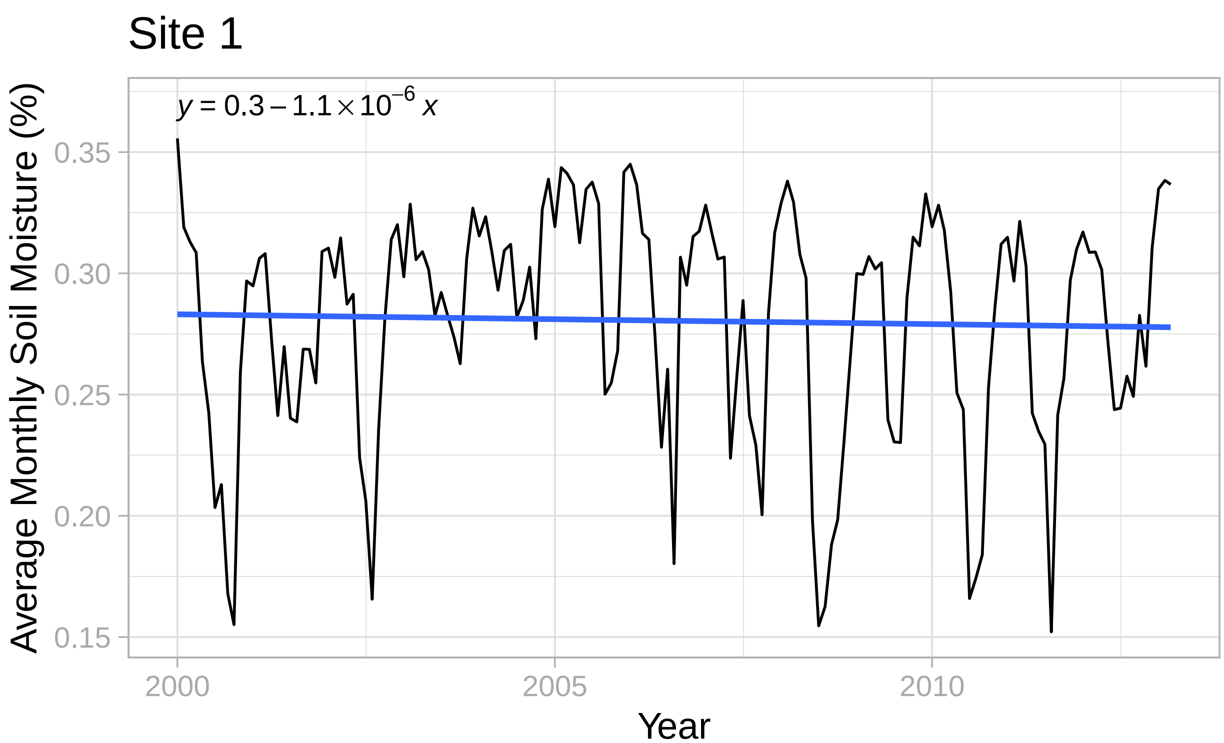
Processed Rain Gauge/Precipitation Data |Variable | Description | Units | Class | Stats |——-:|:————-:|——:|:—–:|:—–:| |Year| Calendar year | | Integer | Minimum = 2000, Maximum = 2013| |Month| Calendar Month | | Integer | Minimum = 01 (January), Maximum = 07 (August)| |AverageMonthlyPrecip| Average Monthly Precipitation | Inches | Numeric | RG5: minimum = 0.1443, mean = 0.6387, maximum = 2.422, RG6: minimum = 0.1033, mean = 0.552, maximum = 1.4345, RG31: minimum = 0.1978, mean = 0.7141, maximum = 2.6956 | |YearMonth| Date in format YYYY-MM-DD | | Character | Minimum = 2000-01-21, Maximum = 2013-07-19|

Variable	**Description**	**Units**	**Class**	**Stats**
Year	Calendar year		Integer	Minimum = 2000 Maximum = 2013
Month	Calendar month		Integer	Minimum = 01 (January) Maximum = 08 (August)
AverageMonthly Precip	Average Monthly Precipitation	Inches	Numeric	**RG5**: Minimum = 0.1443 Mean = 0.6387 Maximum = 2.422
				RG6: Minimum = 0.1033 Mean = 0.552 Maximum = 1.4345
				RG31: Minimum = 0.1978 Mean = 0.7141 Maximum = 2.6956
YearMonth	Date in format YYYY-MM-DD		Character	Minimum = 2000-01-01 Maximum = 2013-08-01

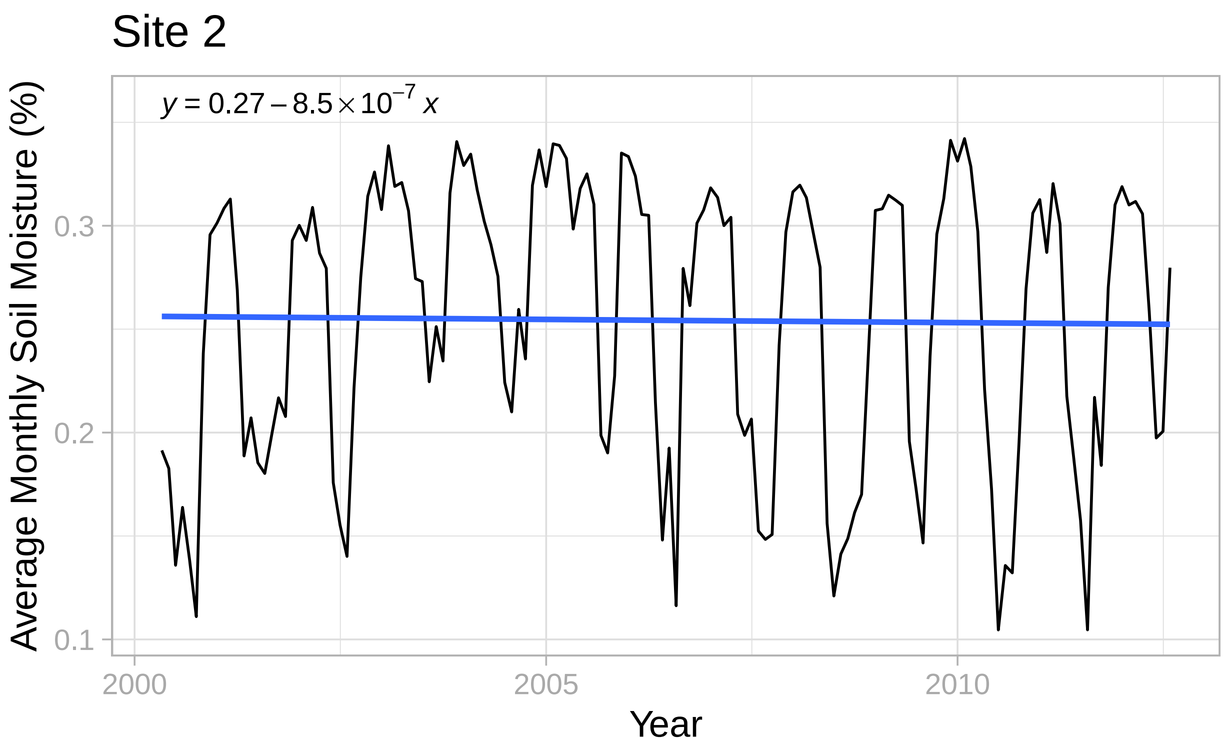
Figure 3: Processed rain gauge/precipitation data.

3 Exploratory Analysis

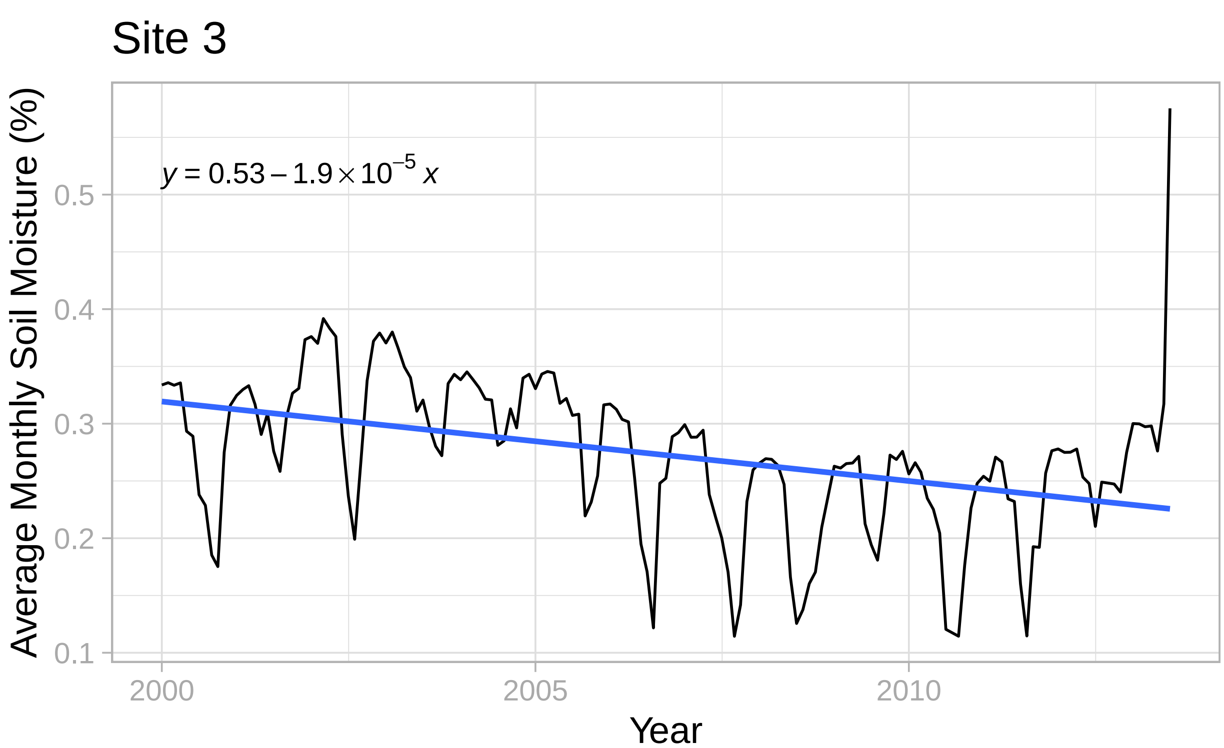
```
## `geom_smooth()` using formula = 'y ~ x'
```



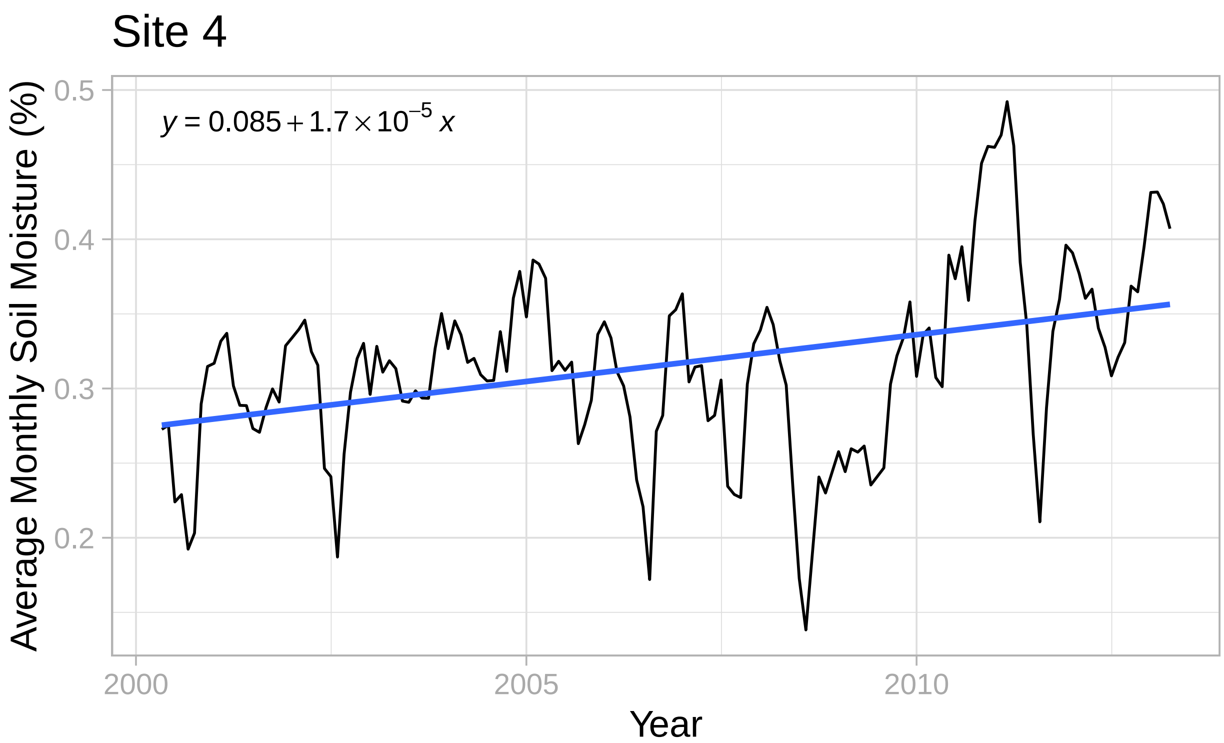

```
## `geom_smooth()` using formula = 'y ~ x'
```



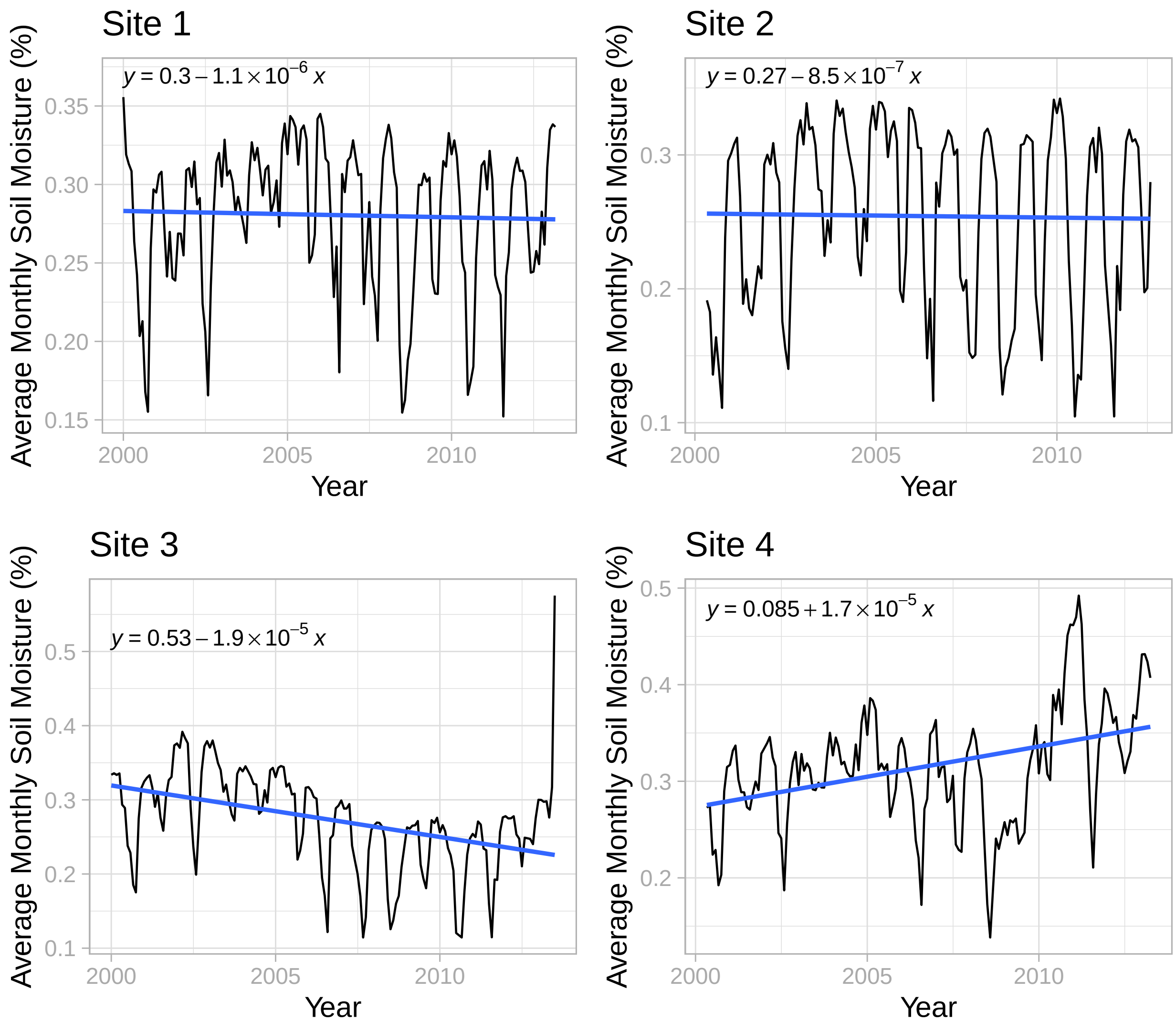

```
## `geom_smooth()` using formula = 'y ~ x'
```



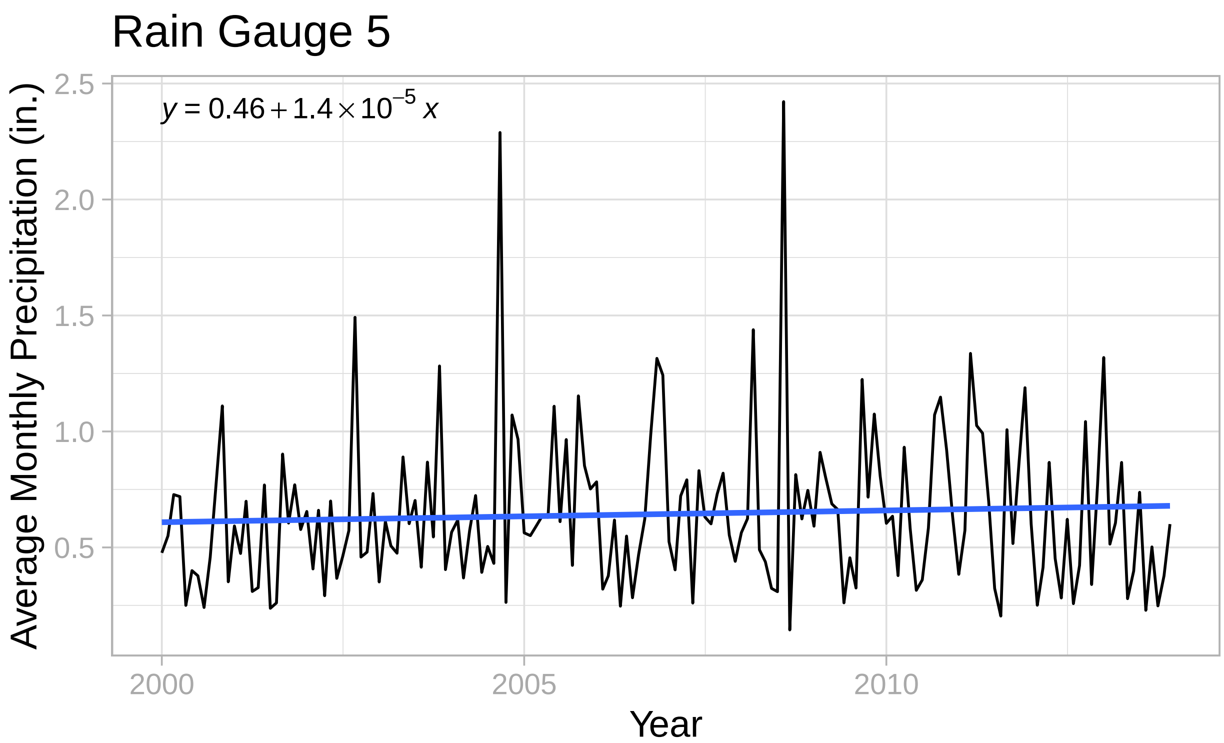

```
## `geom_smooth()` using formula = 'y ~ x'
```



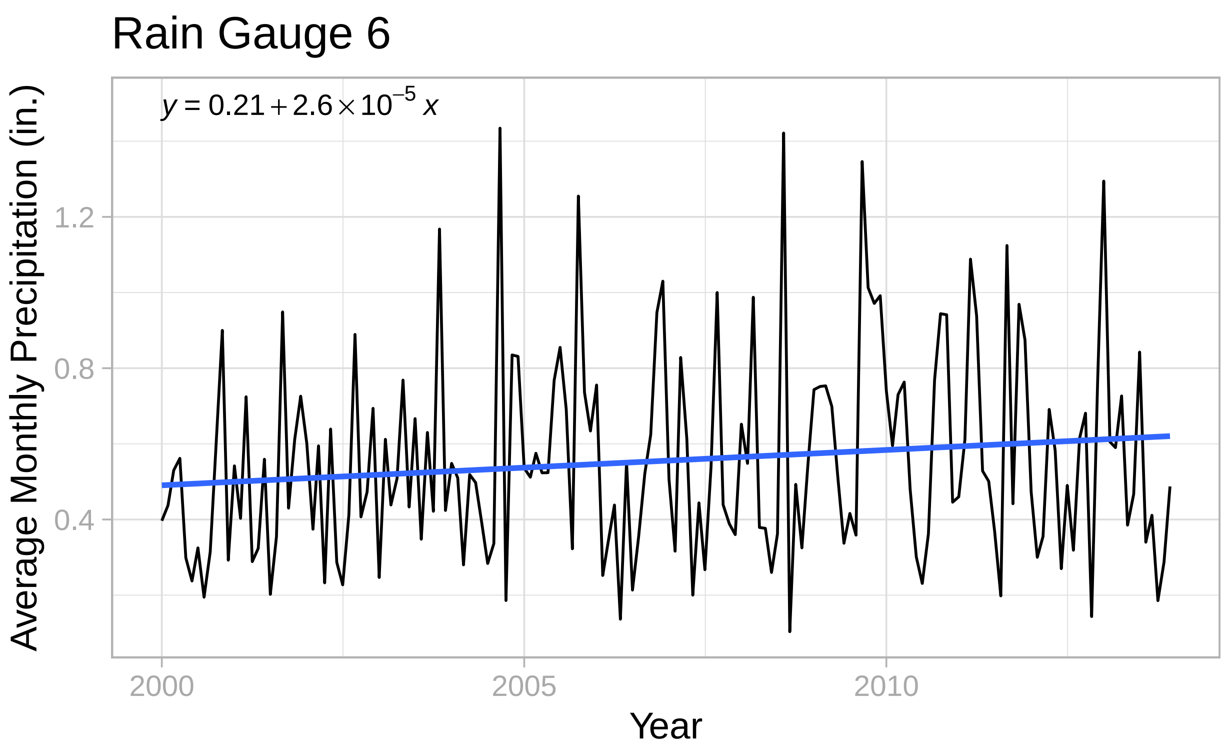

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



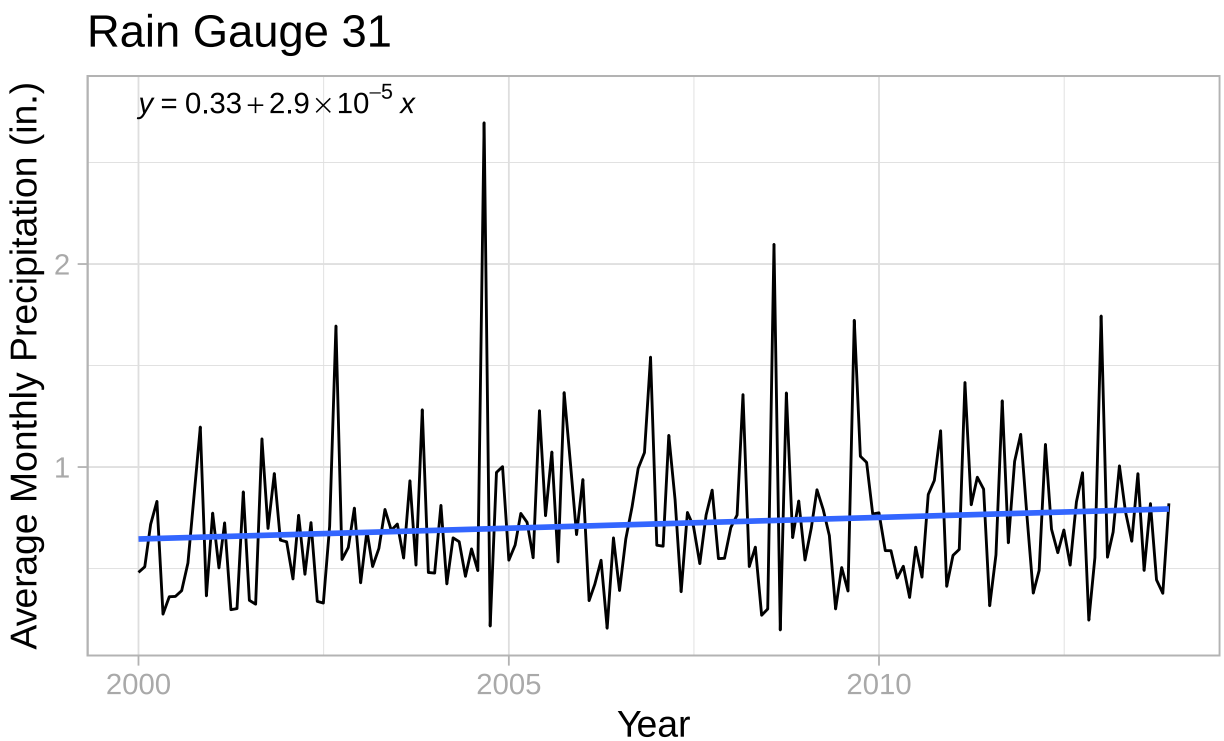

```
## `geom_smooth()` using formula = 'y ~ x'
```



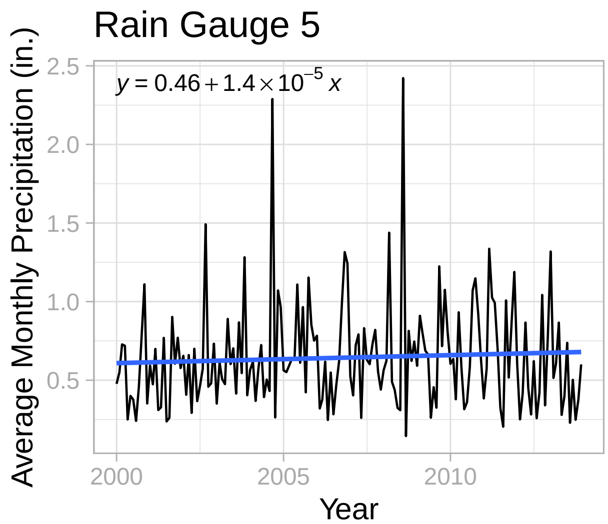

```
## `geom_smooth()` using formula = 'y ~ x'
```

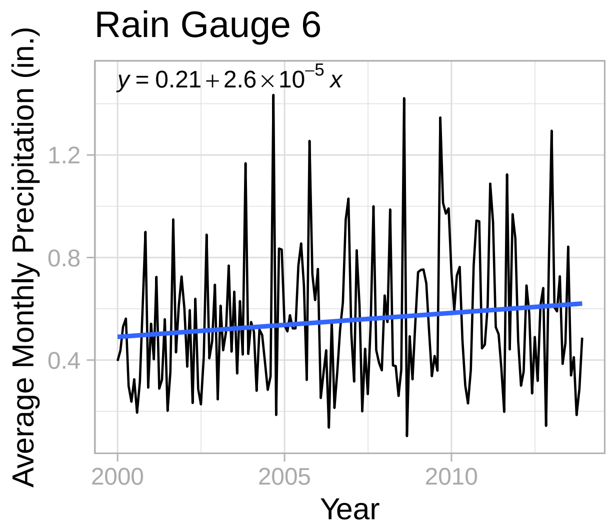


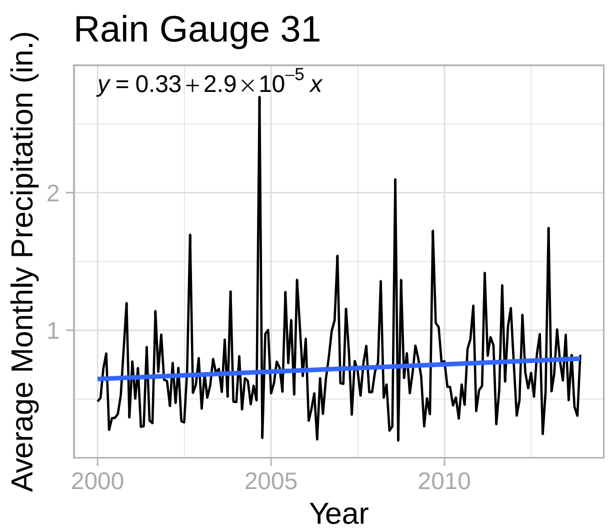

```
## `geom_smooth()` using formula = 'y ~ x'
```



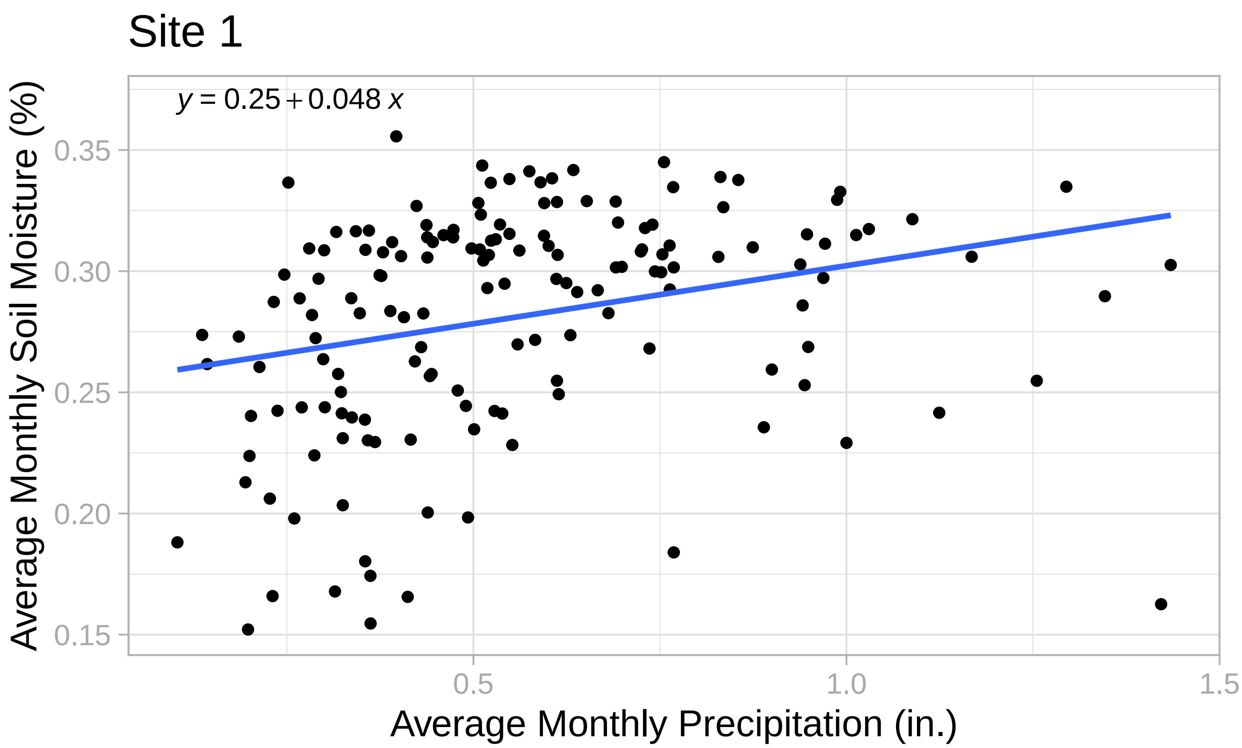

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



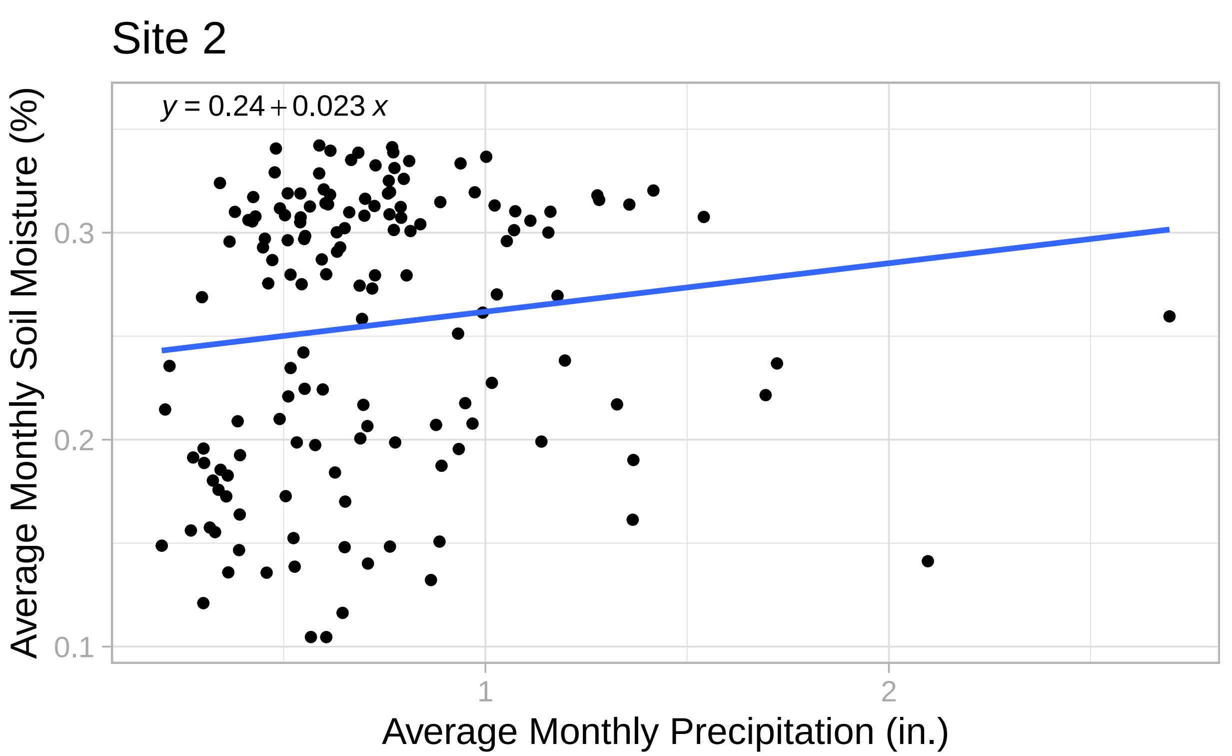




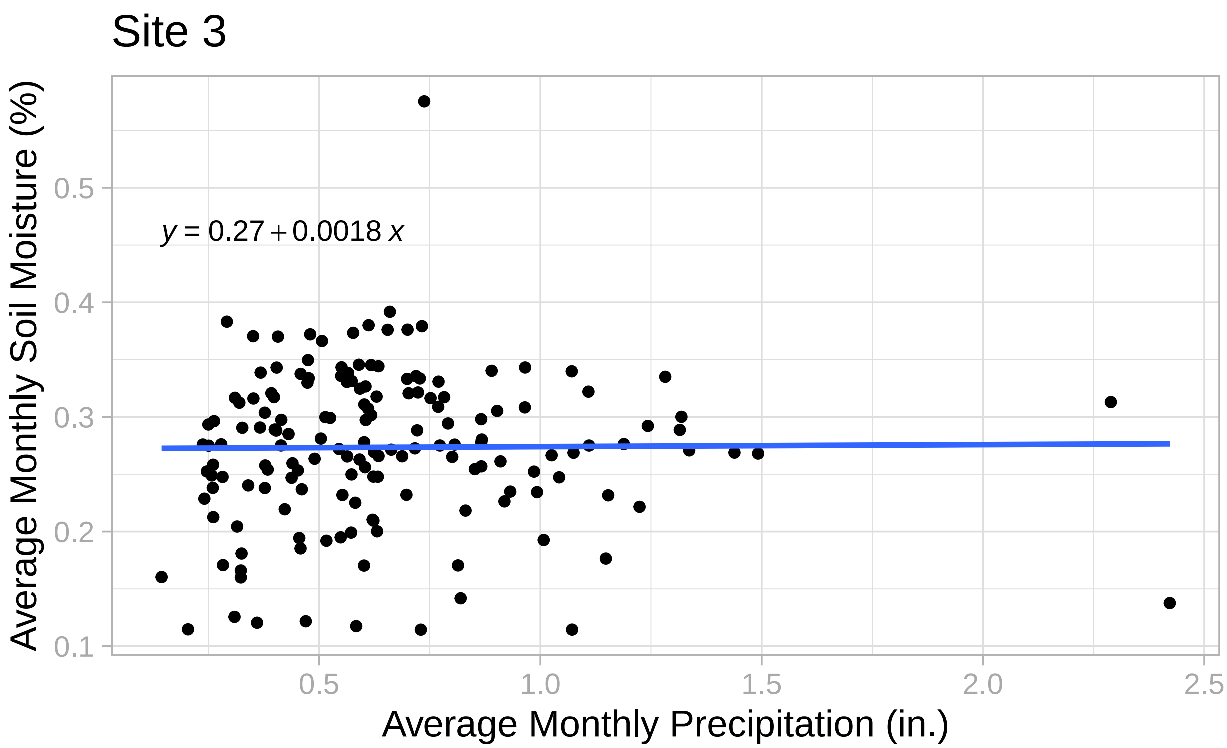

```
## `geom_smooth()` using formula = 'y ~ x'
```



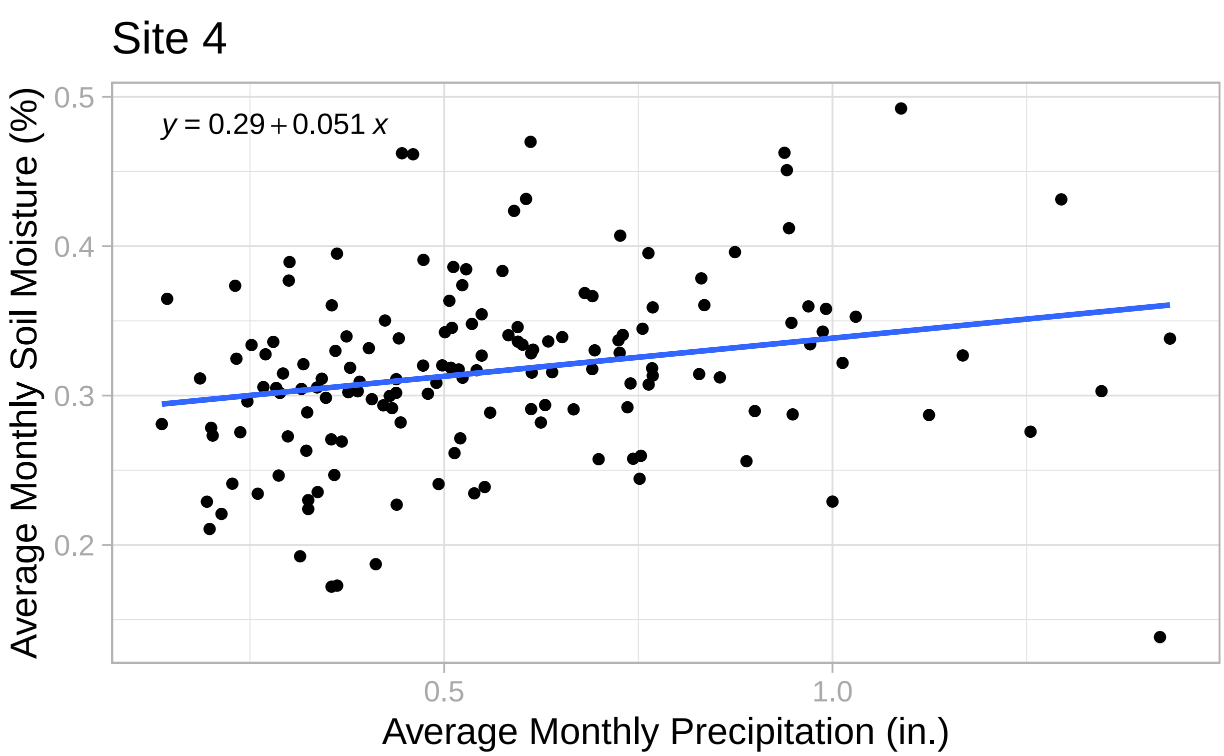

```
## `geom_smooth()` using formula = 'y ~ x'
```




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




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



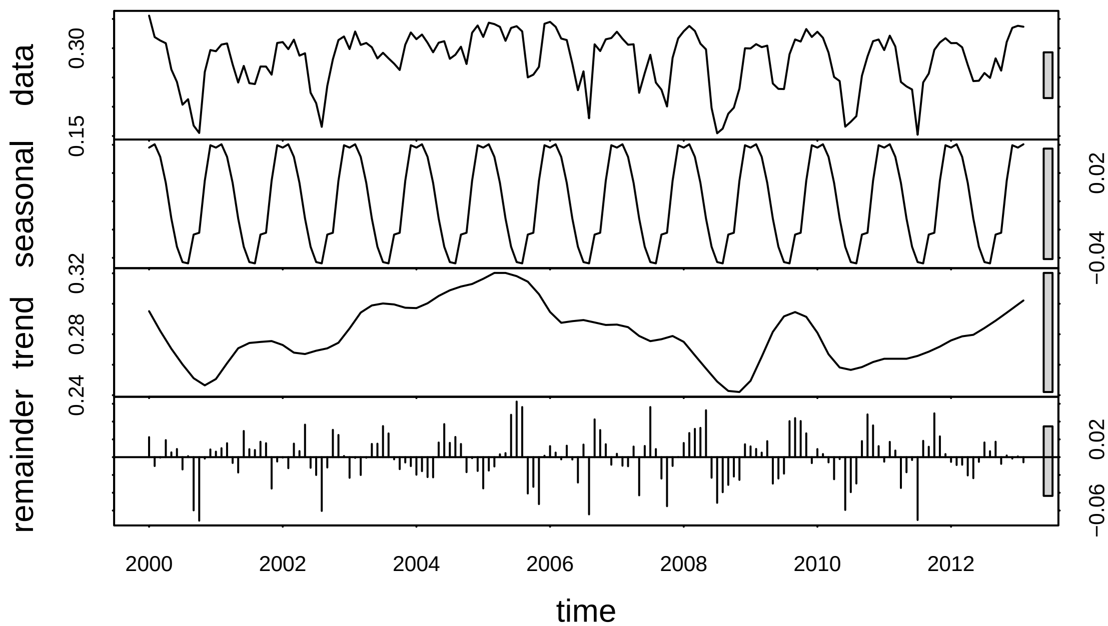
4 Analysis

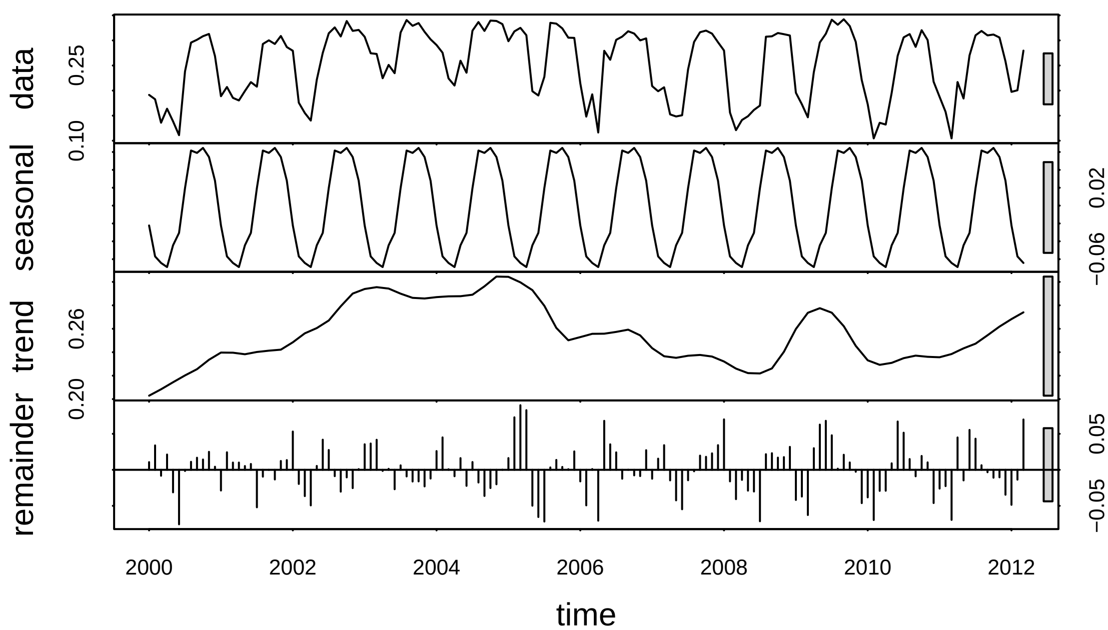
4.1 Question 1: Has soil moisture in the upper 30 centimeters of soil increased or decreased in Coweeta Basin from 2000 to 2013 at each of the four research sites?

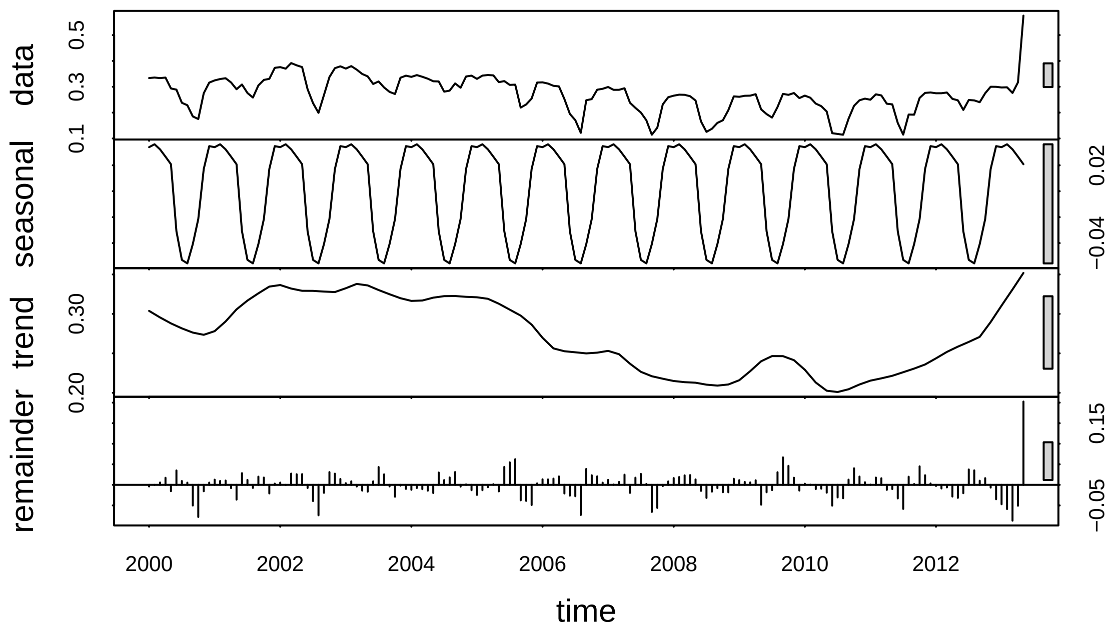
4.2 Question 2: Which sites experienced the greatest change in soil moisture over time?

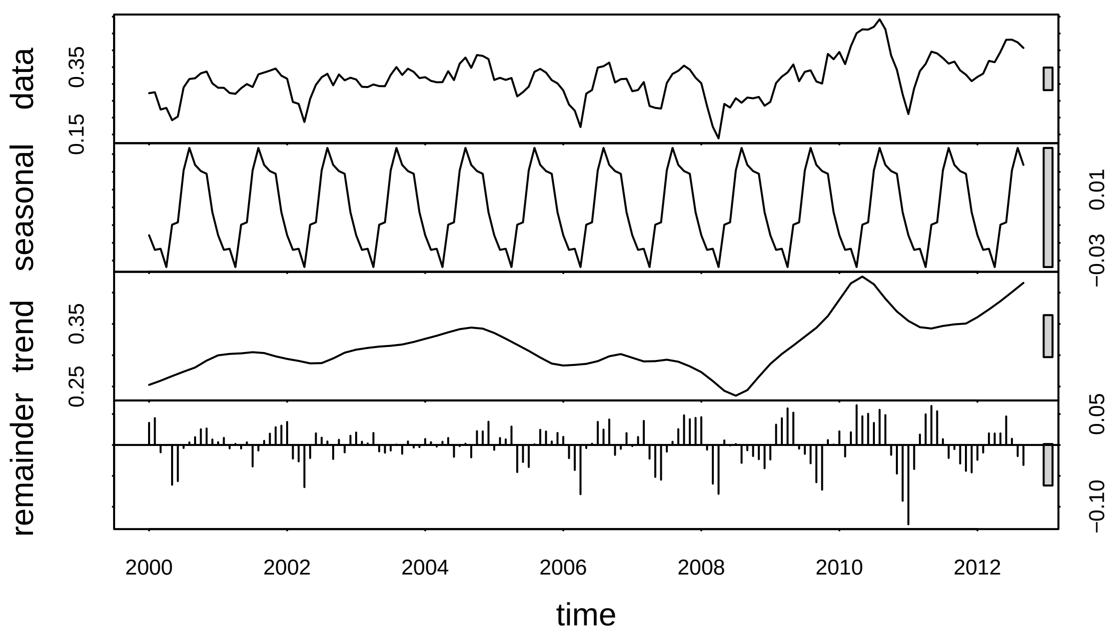
4.3 Question 3: What is the impact of elevation on soil moisture?

In order to analyze potential changes in soil moisture in the upper 30 cm in Coweeta Basin from 2000 to 2013 at each of the four research sites, we completed a time series analysis (TSA) of monthly soil moisture for each site. TSA is used to track a response variable over time to understand whether an increasing or decreasing trend exists across a given time period. Our time series model tracked time in months as the explanatory variable and soil moisture expressed as percent water content as the response variable. TSA can be broken down into the following four components: (1) seasonal component, (2) trend component, (3) error or random component, and (4) cyclical component. For the purpose of our analysis, we focused on the seasonal and trend components in order to interpret general tendencies and repeating cycles, i.e. seasonality. After decomposing and plotting the time series, we compared the length of the gray bars to the right of each plot that explain the relative scale of each component. A longer bar corresponds to smaller variability of that component in the dataset whereas a smaller bar corresponds to larger variability of that component in the dataset. As a result, the trend component explained most of the variability in the dataset of sites 3 and 4 and the seasonality component explained most of the variability in the dataset of sites 1 and 2.


data
seasonal
trend
remainder
0.30
0.15
0.32
0.28
0.24
0.02
−0.04
0.02
−0.06
2000
2002
2004
2006
2008
2010
2012
time


data
seasonal
trend
remainder
0.25
0.10
0.26
0.20
0.02
−0.06
0.05
−0.05
2000
2002
2004
2006
2008
2010
2012
time





We also conducted a Seasonal Mann-Kendall trend analysis of monthly soil moisture for each research site in order to identify seasonal trends in the data. The Seasonal Mann-Kendall trend analysis is a type of monotonic trend analysis test that is non-parametric and can be used on seasonal data that exhibits a repeating pattern at regular intervals over time. A monotonic trend is defined as a gradual change over time that is uniform in direction. The results of the Seasonal Mann-Kendall are as follows:

- *Site 1*: p-value = 0.80898
- *Site 2*: p-value = 0.35739
- *Site 3*: p-value = 3.2565e-12
- *Site 4*: p-value = 1.0157e-5

A p-value less than 0.05 means we must reject the null hypothesis that states the data is stationary and conclude that there is a statistically significant trend in average monthly soil moisture. A p-value greater than 0.05 means we fail to reject the null hypothesis that states the data is stationary and conclude there is not a statistically significant trend in average monthly soil moisture. In reference to the p-values above, sites 1 and 2 have a p-value greater than 0.05 which is not statistically significant. However, sites 3 and 4 have a p-value less than 0.05 which means there is a statistically significant trend at these sites.

```r
#Running the Seasonal Mann-Kendall trend analysis of average monthly soil moisture
site1_trend <- Kendall::SeasonalMannKendall(site1_ts)
summary(site1_trend)

site2_trend <- Kendall::SeasonalMannKendall(site2_ts)
summary(site2_trend)

site3_trend <- Kendall::SeasonalMannKendall(site3_ts)
summary(site3_trend)

site4_trend <- Kendall::SeasonalMannKendall(site4_ts)
summary(site4_trend)
```

4.4 Question 4: Is there a relationship between soil moisture and precipitation that can be observed at these four research sites?

Lastly, we utilized a simple linear regression to determine the strength of the statistical relationship between monthly precipitation and soil moisture. Regression is the fitting of a line to a set of data points and a simple linear regression explains the linear relationship between two variables. The two outputs of the linear model that we chose to focus on were the p-value and multiple R-squared, or coefficient of determination, which is a measure of the percentage of variability in the values of average monthly precipitation (response variable) that is explained by average monthly soil moisture (explanatory variable) and ranges from 0 to 1. The results of the simple linear regression are as follows:

- *Site 1*:
  - p-value = 0.0003247
  - Multiple R-squared = 0.08024
- *Site 2*:
  - p-value = 0.127

    - Multiple R-squared = 0.0161
- *Site 3*:
    - p-value = 0.9118
    - Multiple R-squared = 7.783e-5
- *Site 4*:
    - p-value = 0.00392
    - Multiple R-squared = 0.05413

A p-value less than 0.05 means we must reject the null hypothesis that assumes there is no relationship or correlation between average monthly soil moisture and precipitation and conclude there is a statistically significant relationship between the two variables. A p-value greater than 0.05 means we fail to reject the null hypothesis and conclude there is not a statistically significant relationship between the two variables. In reference to the p-values above, sites 1 and 4 have a significant relationship between average monthly soil moisture and precipitation while sites 2 and 3 do not have a significant relationship between the two variables. Additionally, site 1 had the highest R-squared value of 0.08024 which indicates that approximately 8.02% of the variability in average monthly soil moisture is explained by average monthly precipitation, followed by site 4, site 2, and site 3 in decreasing value. Reviewing the slopes of the linear model for each site also emphasizes the strongest relationship between soil moisture and precipitation at sites 1 and 4 with soil moisture expected to increase approximately 4.79% for every 1 inch increase in precipitation at site 1 and approximately 5.11% for every 1 inch increase in precipitation at site 4.

```
#Running a linear model to determine the strength of the statistical relationship betw
lm_site1 <- lm(data = site1_soil_precip, formula = AverageMonthlySmois30 ~ AverageMonthl
summary(lm_site1)

lm_site2 <- lm(data = site2_soil_precip, formula = AverageMonthlySmois30 ~ AverageMonthl
summary(lm_site2)

lm_site3 <- lm(data = site3_soil_precip, formula = AverageMonthlySmois30 ~ AverageMonthl
summary(lm_site3)

lm_site4 <- lm(data = site4_soil_precip, formula = AverageMonthlySmois30 ~ AverageMonthl
summary(lm_site4)
```

5 Summary and Conclusions

In analyzing the results of these statistical tests, we are able to draw conclusions on soil moisture trends and determine correlation between soil moisture and precipitation from 2000 to 2013 in Coweeta Basin. Our main research question focused on identifying an increasing or decreasing trend in soil moisture in the upper 30 cm at each of the four research sites from 2000 to 2013. The results of the time series analysis and Seasonal Mann-Kendall test indicate that there is a statistically significant trend at sites 3 and 4, with the trend component explaining much of the variability in the dataset of sites 3 and 4 and the seasonality component explaining much of the variability in the dataset of sites 1 and 2. Viewing the plot of average monthly soil moisture over time also provides insight on these trends. Sites 1, 2, and 3 have a decreasing trend in average monthly soil moisture as indicated by a negative slope whereas site 4 has an increasing trend in average monthly soil moisture as indicated by the positive slope. In comparing these trends to determine which sites experienced the greatest change in soil moisture over time, the results of the Seasonal Mann-Kendall test conclude that sites 3 and 4 have the strongest trend in average monthly soil moisture. However, the directionality of these trends differ with a decreasing trend at site 3 and an increasing trend at site 4.

Additional questions that we explored include the impact of elevation on soil moisture as well as the relationship between soil moisture and precipitation at each of the four research sites. Sites 2 and 3 had the highest elevations at approximately 1380 meters and 1215 meters, respectively. Sites 1 and 4 had the lowest elevations at approximately 817 meters and 735 meters, respectively. The results of the Seasonal Mann-Kendall test emphasize a significant trend present at site 4 which is the lowest elevation of all four sites. While an increasing trend of average monthly soil moisture exists at this site, the test also determines that there is not a significant trend at sites 1 or 2. As a result, we are unable to definitively conclude the impact of elevation on soil moisture. In order to address the relationship between soil moisture and precipitation, the results of the linear regression indicated that sites 1 and 4 have a statistically significant relationship between average monthly soil moisture and precipitation. Site 1 had the strongest relationship between the two variables as indicated by the highest R-squared value while site 3 had the weakest relationship as indicated by the smallest R-squared value. In reviewing these conclusions, we do acknowledge that the location of rain gauges and soil moisture sites paired in this study may be a possible limitation or shortcoming. Although the pairing of soil moisture sites and rain gauge stations was based on closest proximity via geographic coordinates, there is uncertainty of local topography in these regions that may lead to slight distortions in the analysis of the relationship between average monthly soil moisture and precipitation.

6 References

<add references here if relevant, otherwise delete this section>