Capstone Project Analysis Code

Introduction to Data Science

Kevin Tajkowski

Library Calls

Load the necessary packages.

```r
library(readr)
library(dplyr)
library(ggplot2)
library(lubridate)
```

Data Preparation

Load the data into a data frame. Remove all non-fish species by Family_Name. Non-fish species observed in the data set include sharks, rays, eels, turtles and a dolphin.

```r
Little_Cayman <- read_csv("Little_Cayman.csv")
non_fish_family <- c("Requiem Shark", "Conger", "Dolphin", "Stingray",
"Manta", "Moray", "Eagle Ray", "Snake Eel", "Carpet Shark", "Hammerhead
Shark", "Electric Ray", "Round Stingrays", "Sea Turtles", "#N/A")
LC <- Little_Cayman %>% filter(!(Family_Name %in% non_fish_family))
```

Data Frame Manipulation

Red lionfish are removed from the data set. Fish families with less than a prescribed number of surveys for all species in the family are identified and removed from the data set.

Density Comparison

'LC_Mod' is already created (excludes non-fish species, lionfish and fish families with < a certain number of surveys) 'family_list' is a vector of family names already created

Generate a density plot of the slopes of the trendlines pre and post-lionfish

Pre and Post-Lionfish Density Comparison
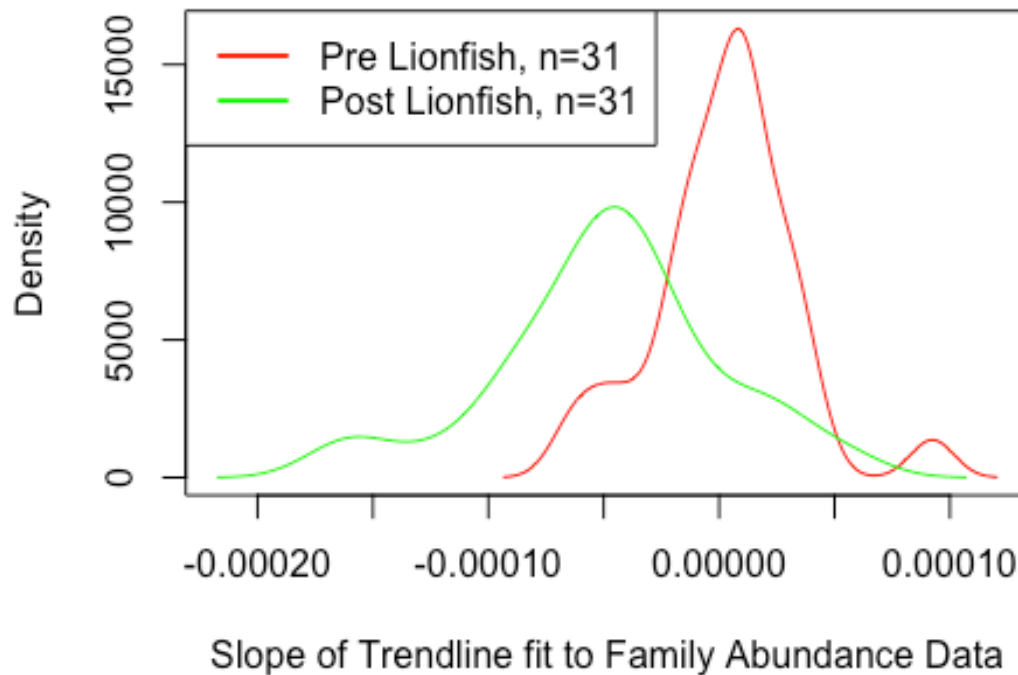
Moving Window Density Comparison

Use a moving window to analyze the data from specific time frames progressing from the first survey date (March 27, 1994) to the last survey date available (April 17, 2018). The size of the window (number of years) and the time to shift the window (number of years) are variables.

The following code is used to identify important dates and initialize several variables.

```
initial_date <- as.Date("1994-03-27") # Date that the first survey was
taken
final_date <- as.Date("2018-04-17") # Date of the last survey taken
shift <- 1 # Duration from beginning of window to the beginning of next
window (in years)
window <- 2 # Size (length of time) of window (in years) - Should be eq
ual to or larger than 'shift'

begin_date <- initial_date # initialize beginning of window to date of
first survey
end_date <- begin_date %m+% years(window) # initialize end of window
```

```r
mean_slopes <- NA
i <- 1
j <- 1
```

The following code is used to determine the average slope of all of the slopes of the trendlines for every family during a specific time period. The time period corresponds to the moving window.

```r
while(end_date <= final_date){
  current_window <- LC_Mod %>% filter(Date >= begin_date & Date < end_d
ate)
  slopes <- NA

  for(i in seq_along(family_list)){
    c_wind <- current_window %>% filter(Family_Name == family_list[i])
    TL_slope <- lm(c_wind$Abundance ~ c_wind$Date)
    slopes[i] <- TL_slope$coefficients["c_wind$Date"]
  }
  mean_slopes[j] <- mean(slopes)
  j <- j+1

  d1 <- density(slopes)
  xlim <- range(d1$x)
  ylim <- range(d1$y)
  cols <- "red"
  plot(1, xlim=xlim, ylim=ylim, type='n', main=paste(begin_date, "to",
end_date, "\nWindow length:", window, "years"), xlab='Slope of Trendlin
e fit to Family Abundance Data', ylab='Density')
  rug(slopes)
  density(slopes, adjust = 0.5)
  lines(d1$x, d1$y, col=cols)

  begin_date <- begin_date %m+% years(shift)
  end_date <- begin_date %m+% years(window)
}
```
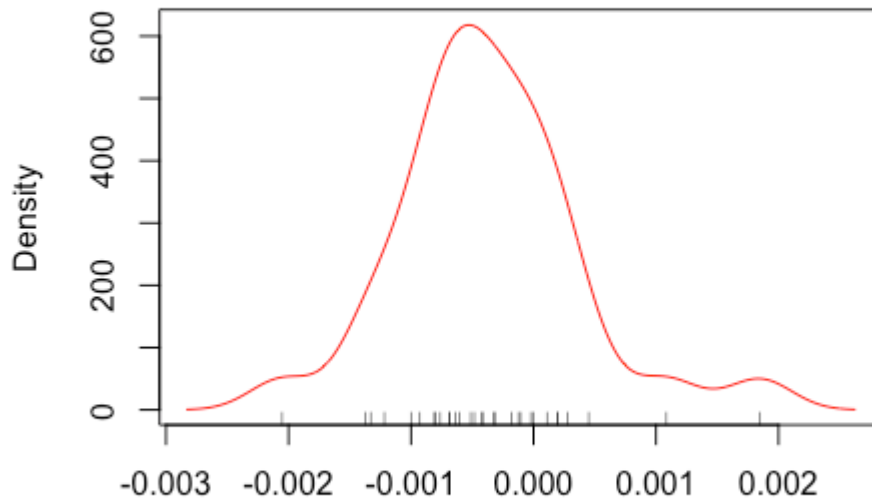
1994-03-27 to 1996-03-27
Window length: 2 years

Density

500
300
100
0

-0.020   -0.015   -0.010   -0.005   0.000

Slope of Trendline fit to Family Abundance Data
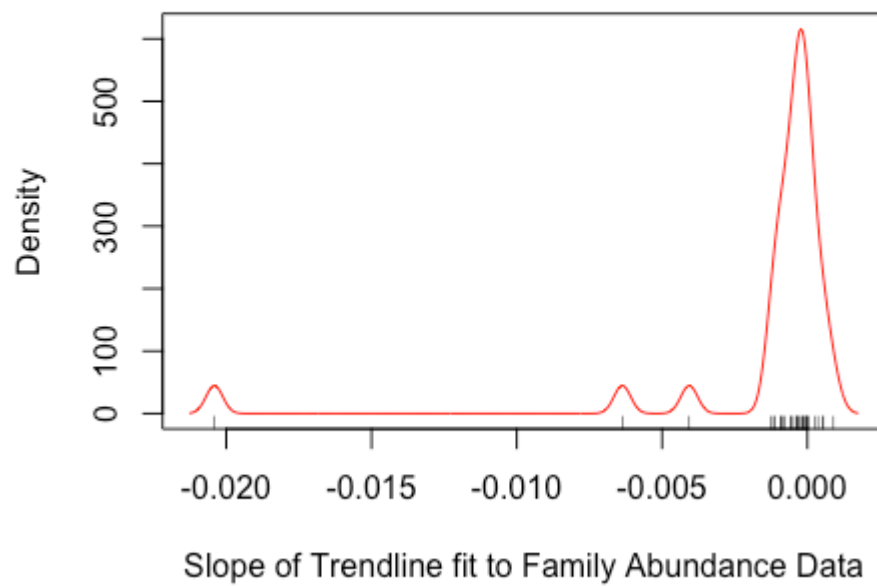
1995-03-27 to 1997-03-27
Window length: 2 years

Density
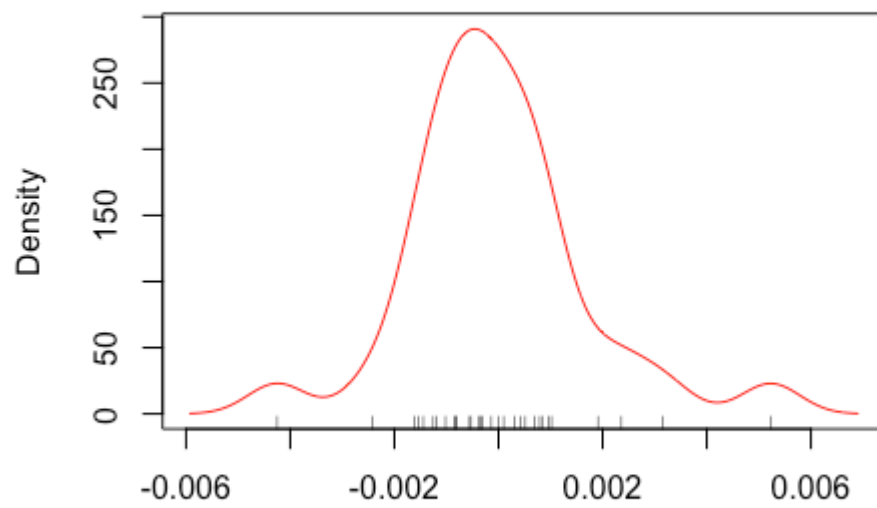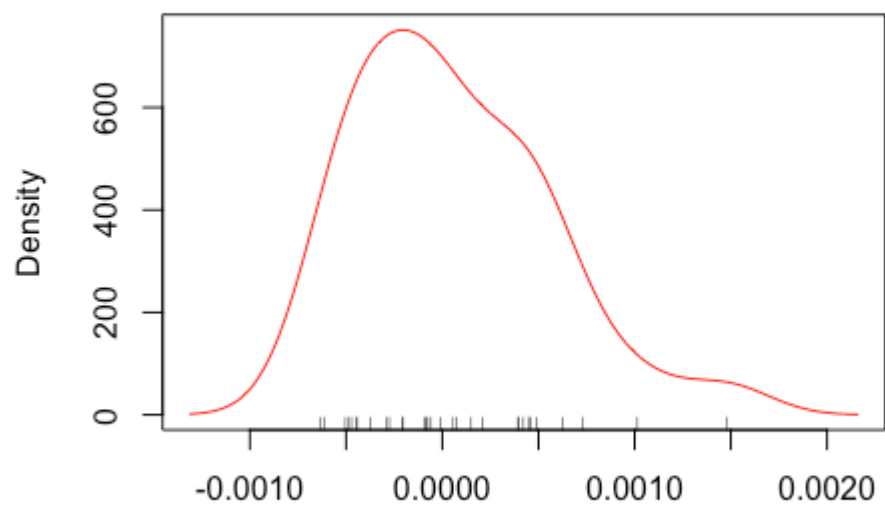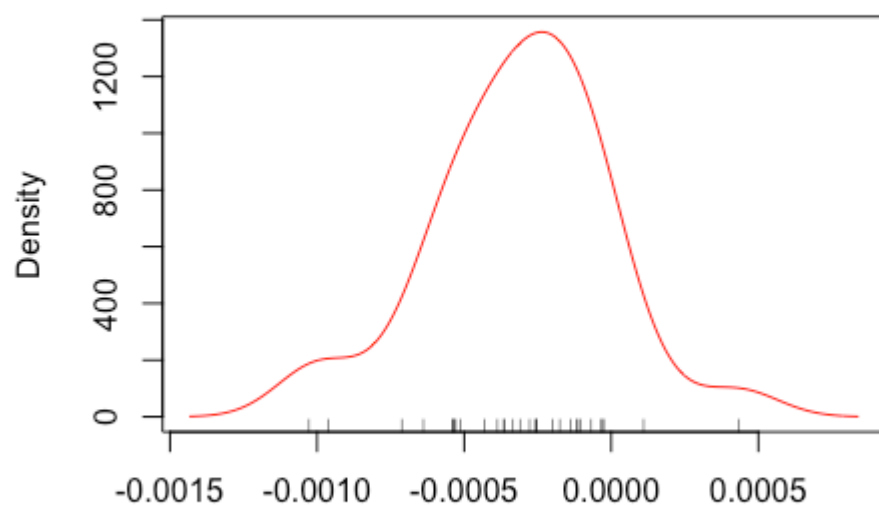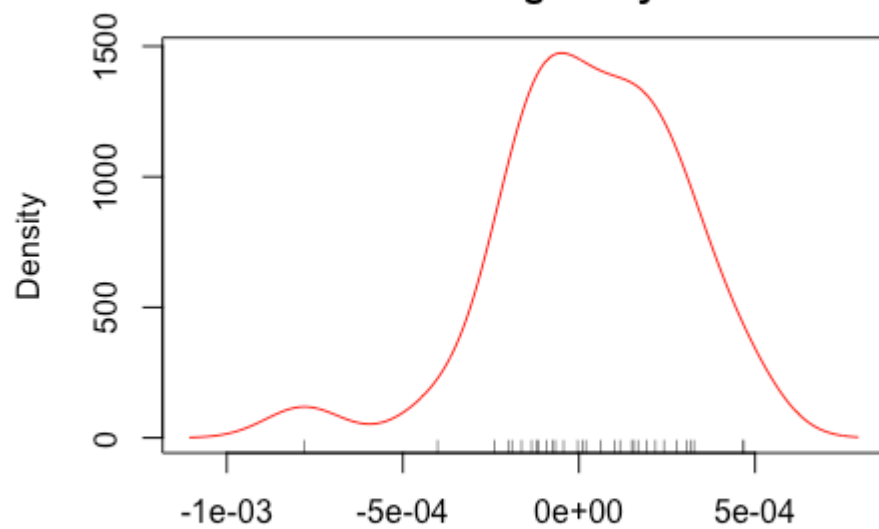
800
600
400
200
0

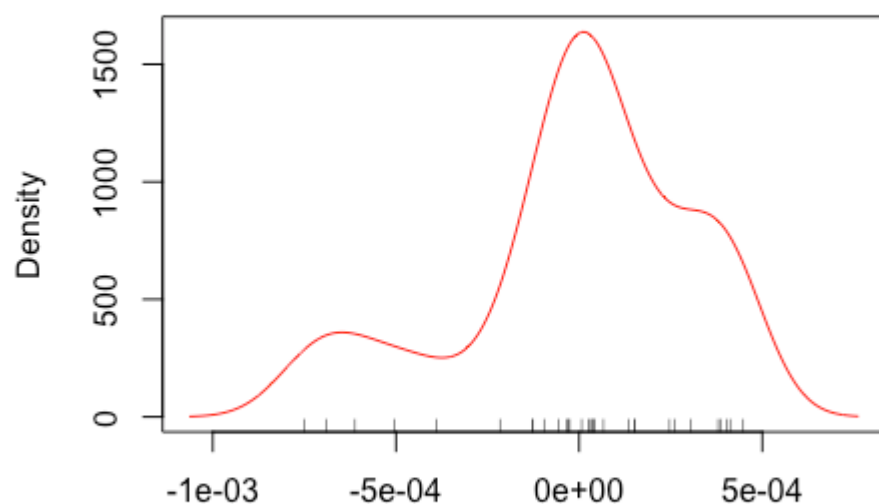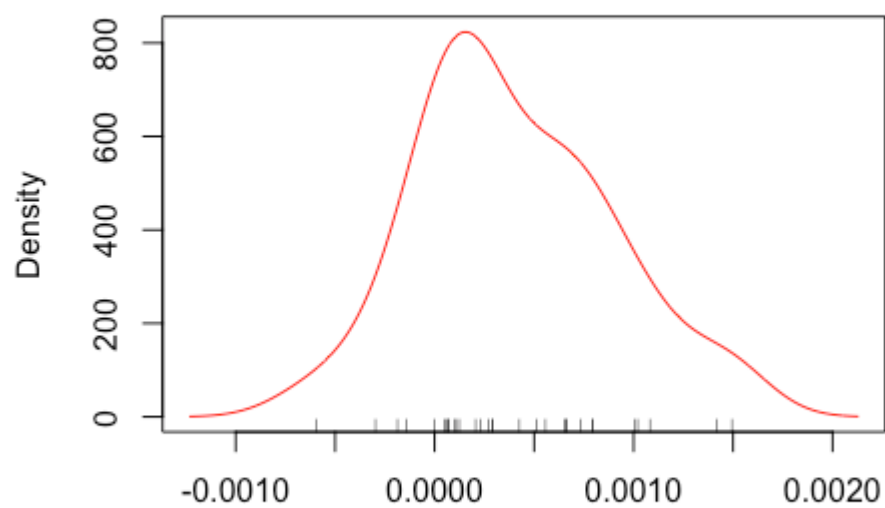-0.001   0.000   0.001   0.002

Slope of Trendline fit to Family Abundance Data
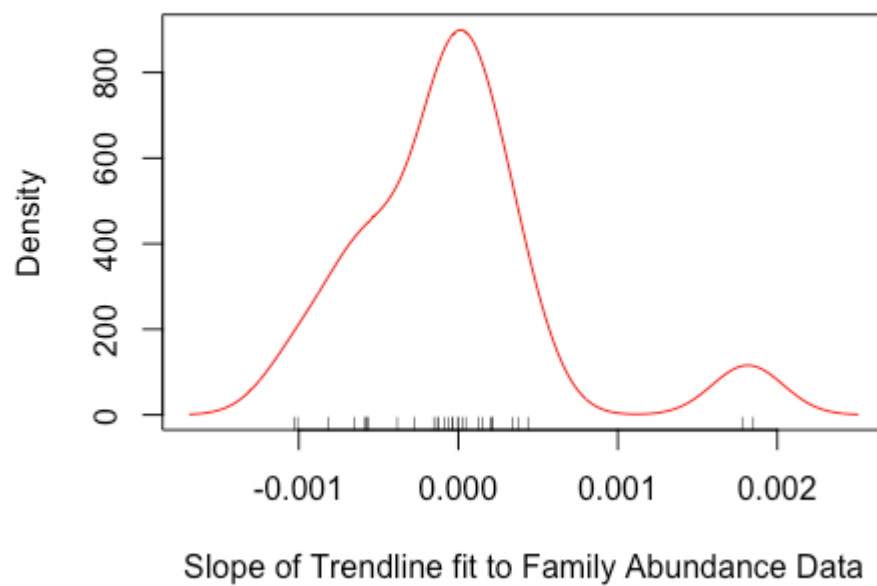
1996-03-27 to 1998-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

1997-03-27 to 1999-03-27
Window length: 2 years
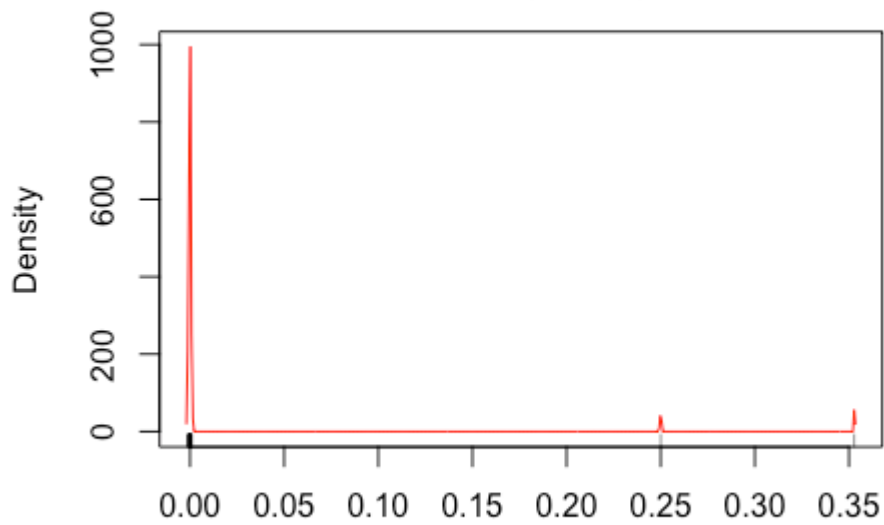
Slope of Trendline fit to Family Abundance Data

1998-03-27 to 2000-03-27
Window length: 2 years

Density

Slope of Trendline fit to Family Abundance Data

1999-03-27 to 2001-03-27
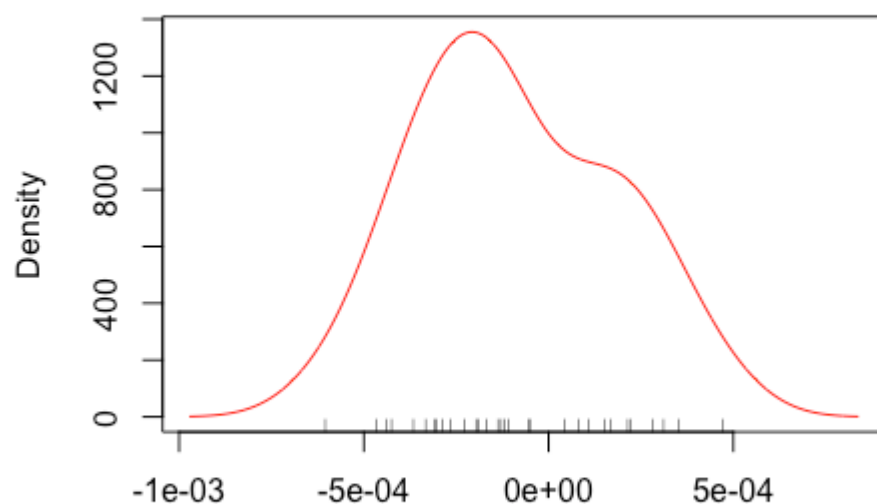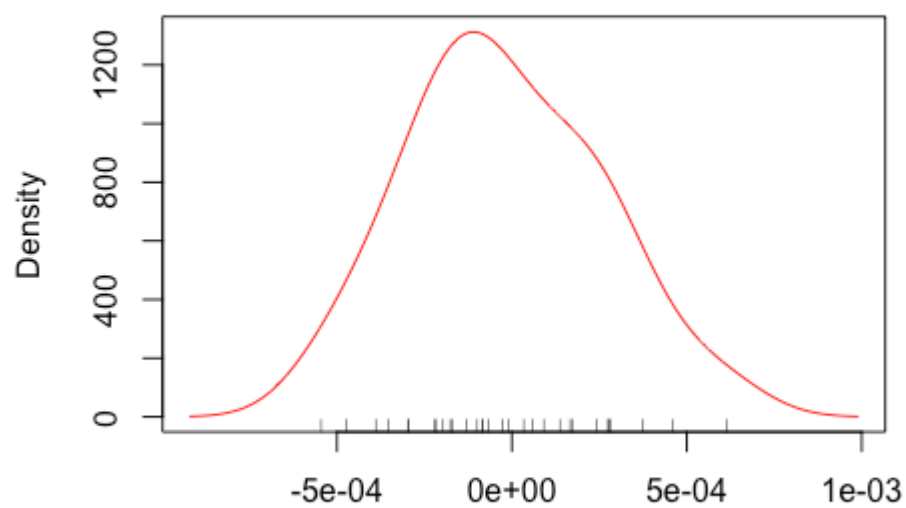## Window length: 2 years

Density

Slope of Trendline fit to Family Abundance Data

2000-03-27 to 2002-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

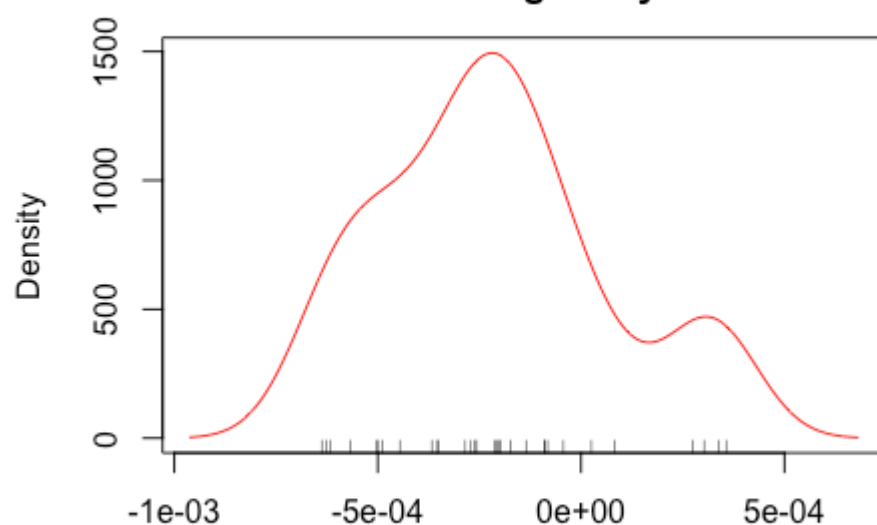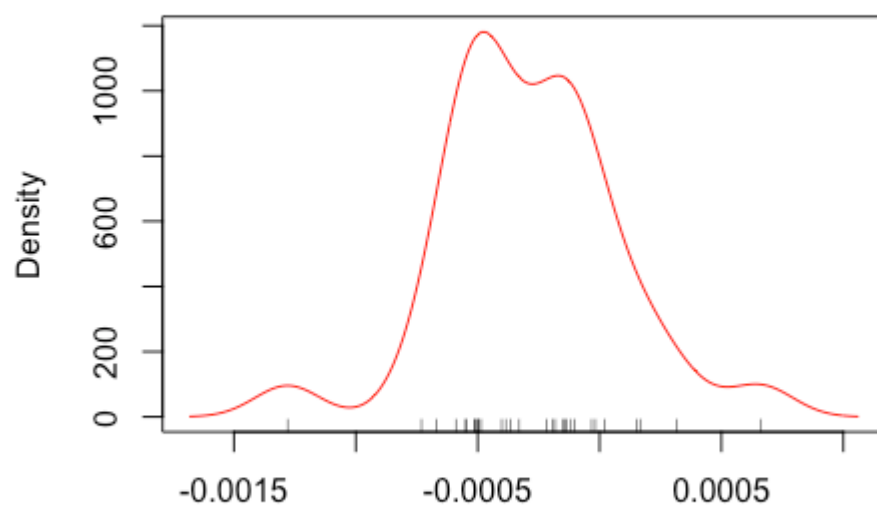## 2001-03-27 to 2003-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2002-03-27 to 2004-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

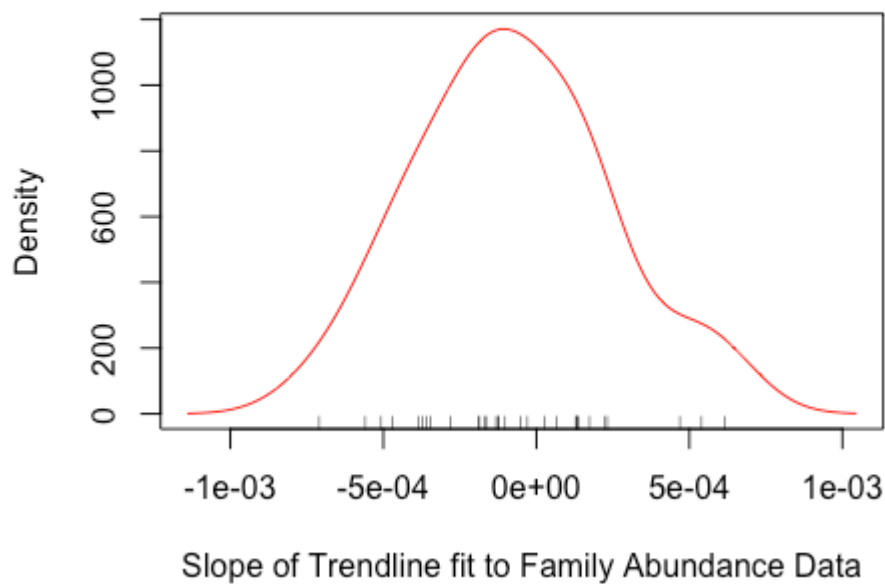## 2003-03-27 to 2005-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2004-03-27 to 2006-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2005-03-27 to 2007-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2006-03-27 to 2008-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data
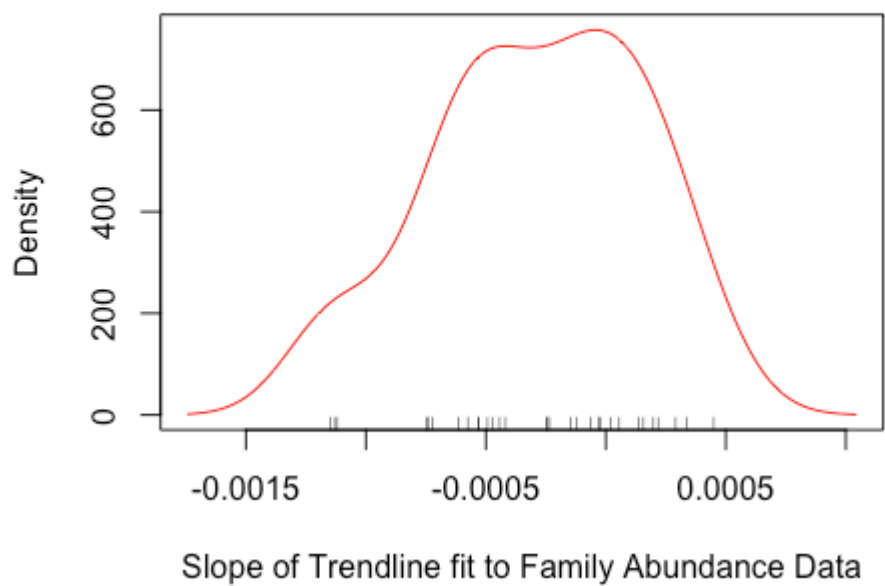
2007-03-27 to 2009-03-27
Window length: 2 years

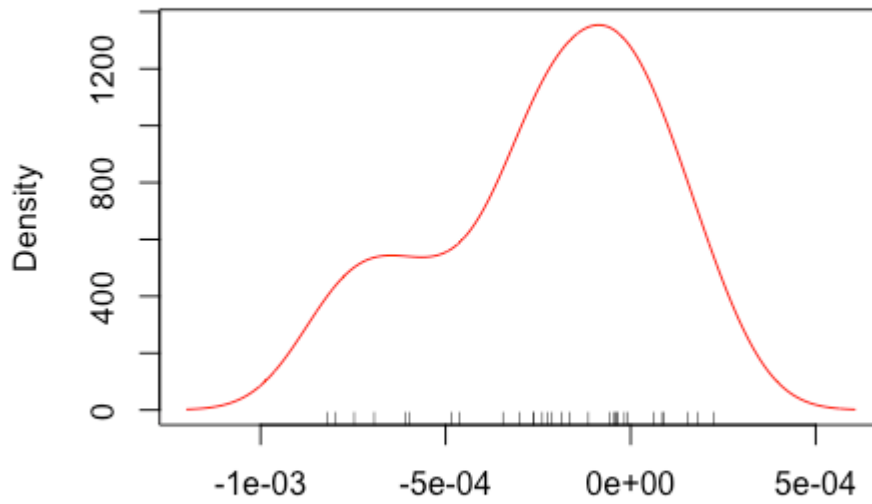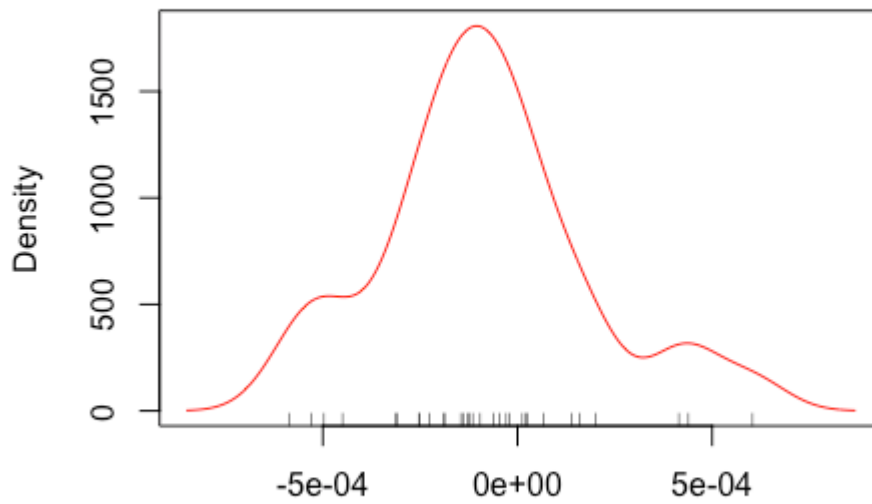Slope of Trendline fit to Family Abundance Data

2008-03-27 to 2010-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2009-03-27 to 2011-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2010-03-27 to 2012-03-27
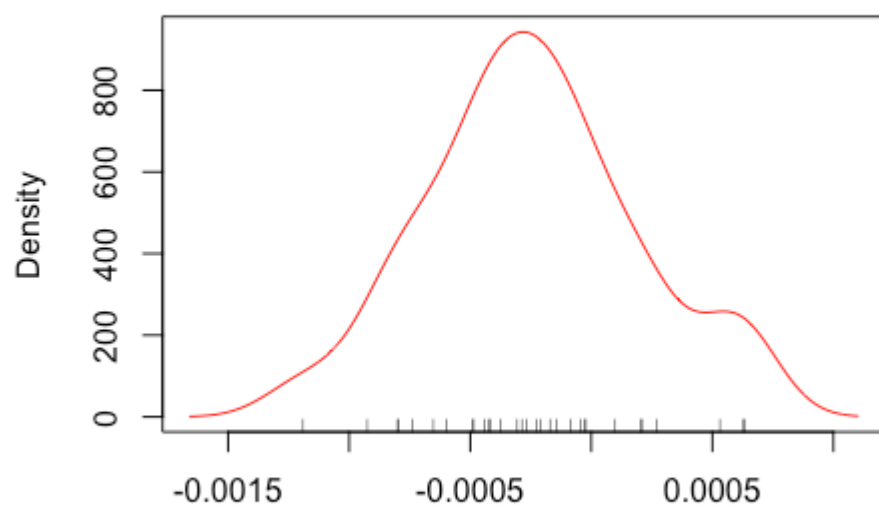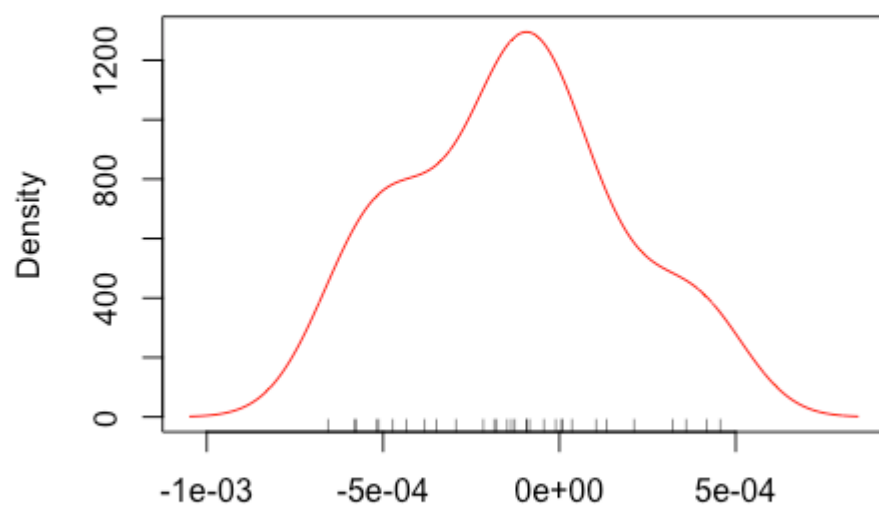**Window length: 2 years**

Slope of Trendline fit to Family Abundance Data

2011-03-27 to 2013-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2012-03-27 to 2014-03-27
Window length: 2 years

Density

Slope of Trendline fit to Family Abundance Data

2013-03-27 to 2015-03-27
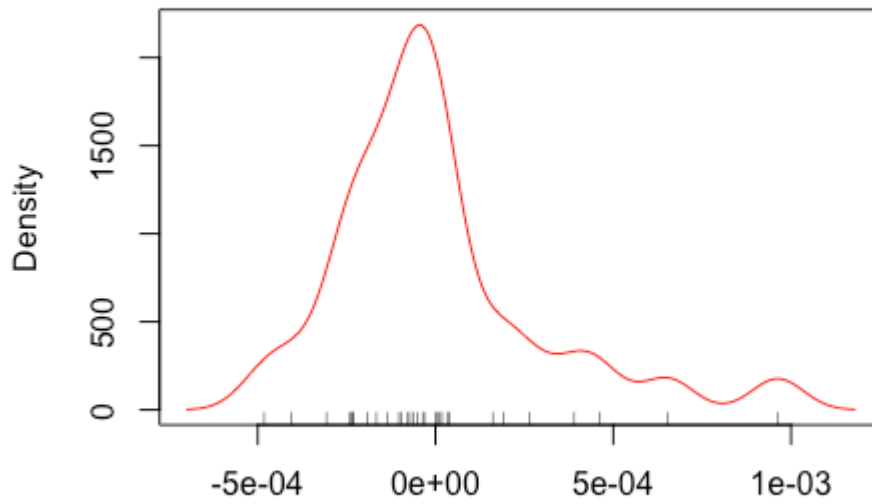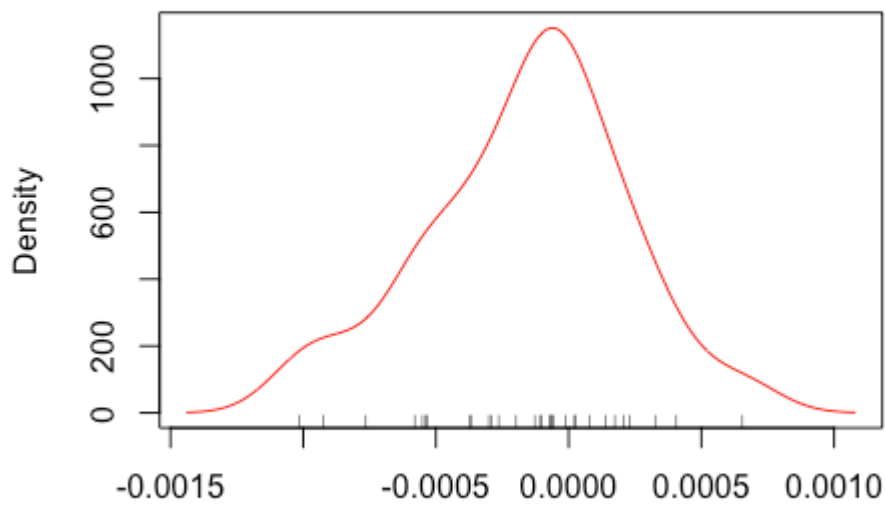**Window length: 2 years**

Density

Slope of Trendline fit to Family Abundance Data

2014-03-27 to 2016-03-27
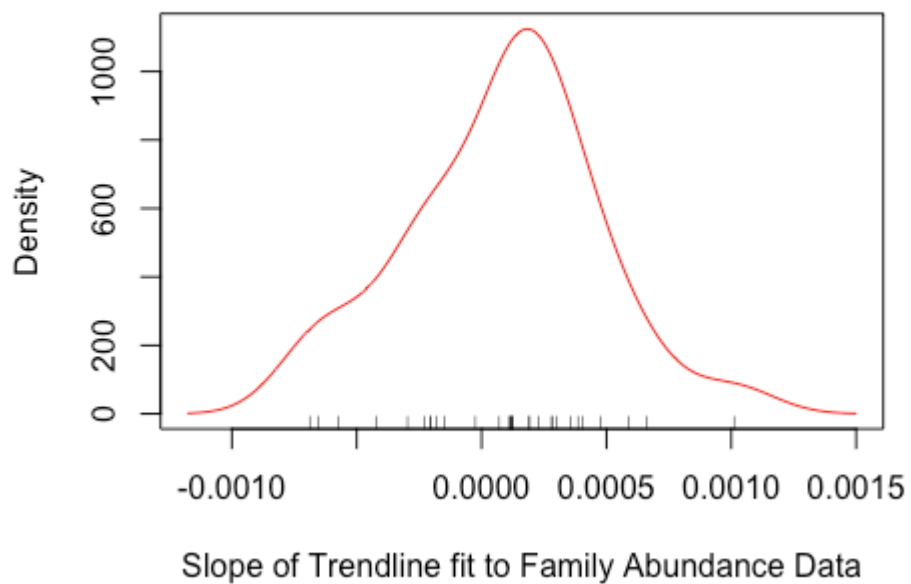**Window length: 2 years**

Slope of Trendline fit to Family Abundance Data

2015-03-27 to 2017-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

2016-03-27 to 2018-03-27
Window length: 2 years

Slope of Trendline fit to Family Abundance Data

```
plot(mean_slopes, main = "Mean Slope of Trendlines for all Families\nfo
r Moving Window of Survey Data", ylab = "Mean Slope")
abline(h = 0,  col = "green")
lines(mean_slopes, type = "l")
```

Mean Slope of Trendlines for all Families for Moving Window of Survey Data