1 Introduction

This appendix provides documentation and R code association with the paleodiatom community example.

1.1 Setup

1.1.1 Load required packages

You will need to install the following packages if they are not already.

```
## -- Attaching packages ----------
```

```
## v ggplot2 3.1.0       v purrr   0.3.2
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0
```

```
## -- Conflicts -------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: ggridges
```

```
##
## Attaching package: 'ggridges'
```

```
## The following object is masked from 'package:ggplot2':
##
##     scale_discrete_manual
```

```
## The ggjoy package has been deprecated. Please switch over to the
## ggridges package, which provides the same functionality. Porting
## guidelines can be found here:
## https://github.com/clauswilke/ggjoy/blob/master/README.md
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

1.1.2 Load the data

Pull in the data from the supplementary materials for Spanbauer *et al.* (2014) (Spanbauer et al. 2014). This data contains the percent abundances of diatom species from Foy Lake. Spanbauer *et al.* (2014) calculated the number of relative diatom valves in each sample. They removed time steps with no diatom data, claiming poor preservation rather than zero abundance. The authors also averages time steps 301 - 312.

```
data <- read_csv("https://doi.org/10.1371/journal.pone.0108936.s001")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Sample = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

1.2 Calculate the distance metric

1.2.1 Calculate the distance

Calculate the difference, Δx, in each species' relative abundance from time n to time $n + 1$. Calculate the change in distance, Δs, as the sum of the squares of the change in each species. Calculate the distance as the cummulative sums of the change in distance.

```
distance <-
  data %>%
  select(-Sample) %>%
  arrange(YB1950) %>%
  mutate_at(funs(lead(.)-.), .vars = vars(-YB1950)) %>%
  gather(species, dx, -YB1950) %>%
  group_by(YB1950) %>%
  summarize(ds = sqrt(sum(dx^2))) %>%
  mutate(s = cumsum(ds)) %>%
  filter(!is.na(s))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## please use list() instead
##
## # Before:
## funs(name = f(.)
##
## # After:
## list(name = ~f(.))
## This warning is displayed once per session.
```

Load the numerical differentiation results data. These finite differences were approximated in MatLab, using code that implements the methods found in Chartrand, R. Numerical differentiation of noisy, nonsmooth data. (2011) ISRN Applied Mathematics, Vol. 2011. Article ID 164564 (Chartrand 2011).

```
# Load csv file
numDiff <- read.csv("numerical_diff_results.csv", header = F)
names(numDiff) <- c("t","dsdt","d2sdt2")

# Create vectors of for plotting purposes
tt <- numDiff$t
vv <- numDiff$dsdt
aa <- numDiff$d2sdt
```

1.2.2 Dataframe processing

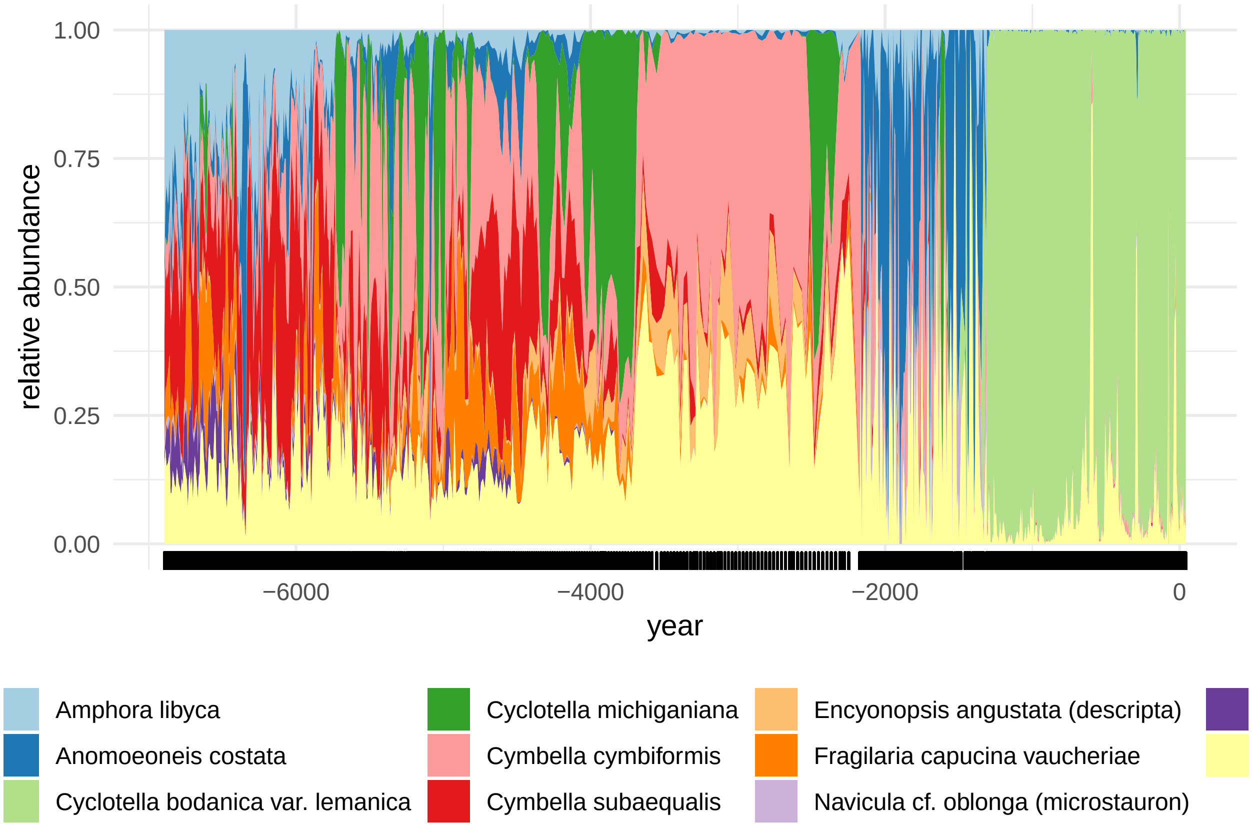
Create a dataset in 'long' form for future plotting

```
longData <-
  data %>%
  select(-Sample) %>%
  gather(species, relAbund, -YB1950)

# Find some species that have "high" abundances
topSpecies <-
  longData %>%
  group_by(species) %>%
  summarize(q95 = quantile(relAbund, 0.95)) %>%
  arrange(desc(q95)) %>%
  filter(row_number() <= 10)
```

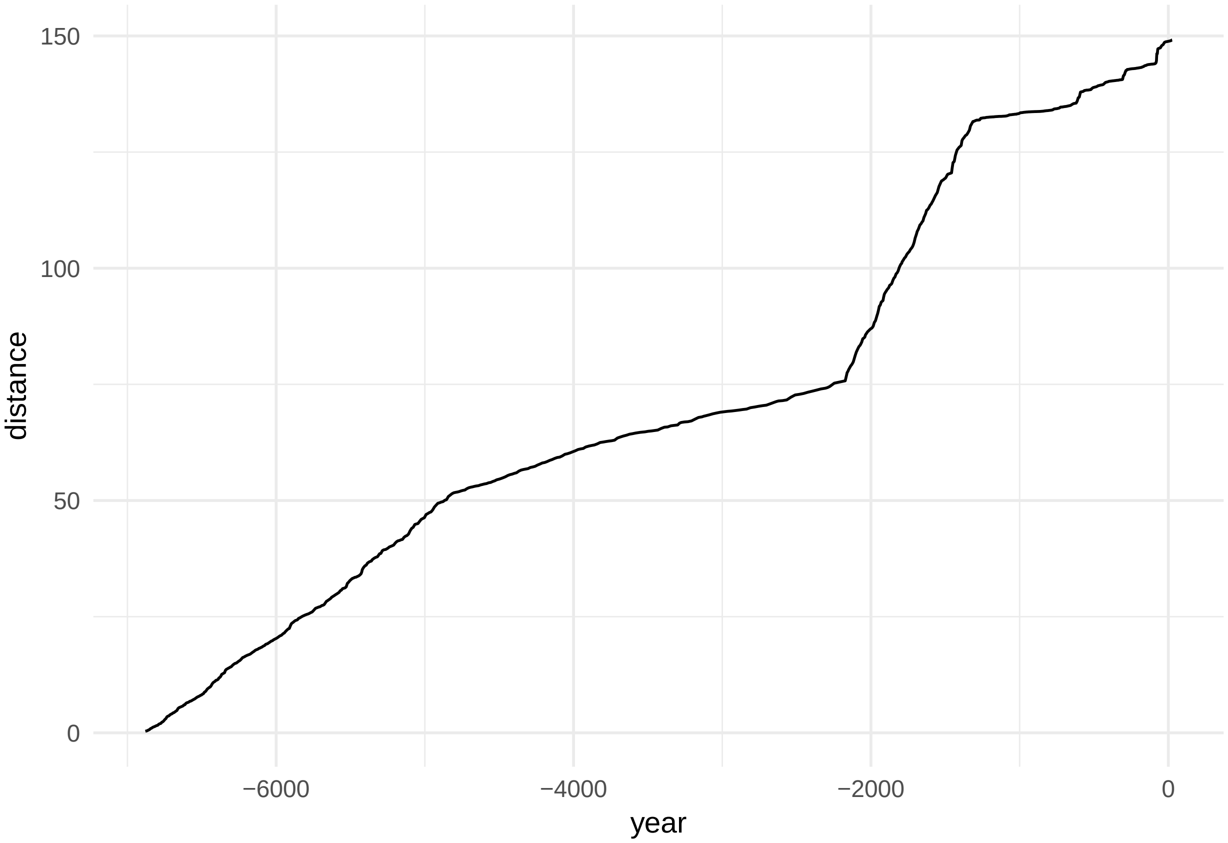
1.3 Plotting

1.3.1 Plot the relative abundance data over time

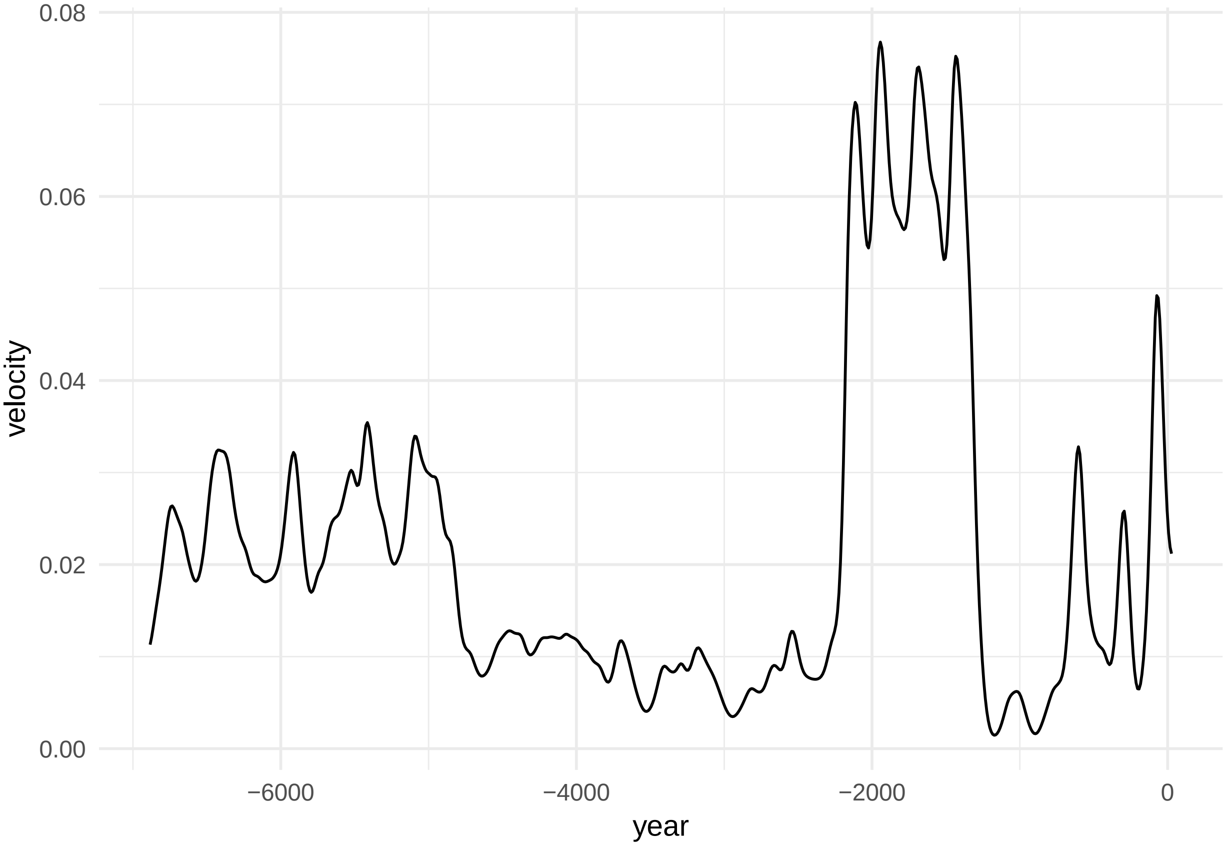


1.3.2 Plot the distance travelled

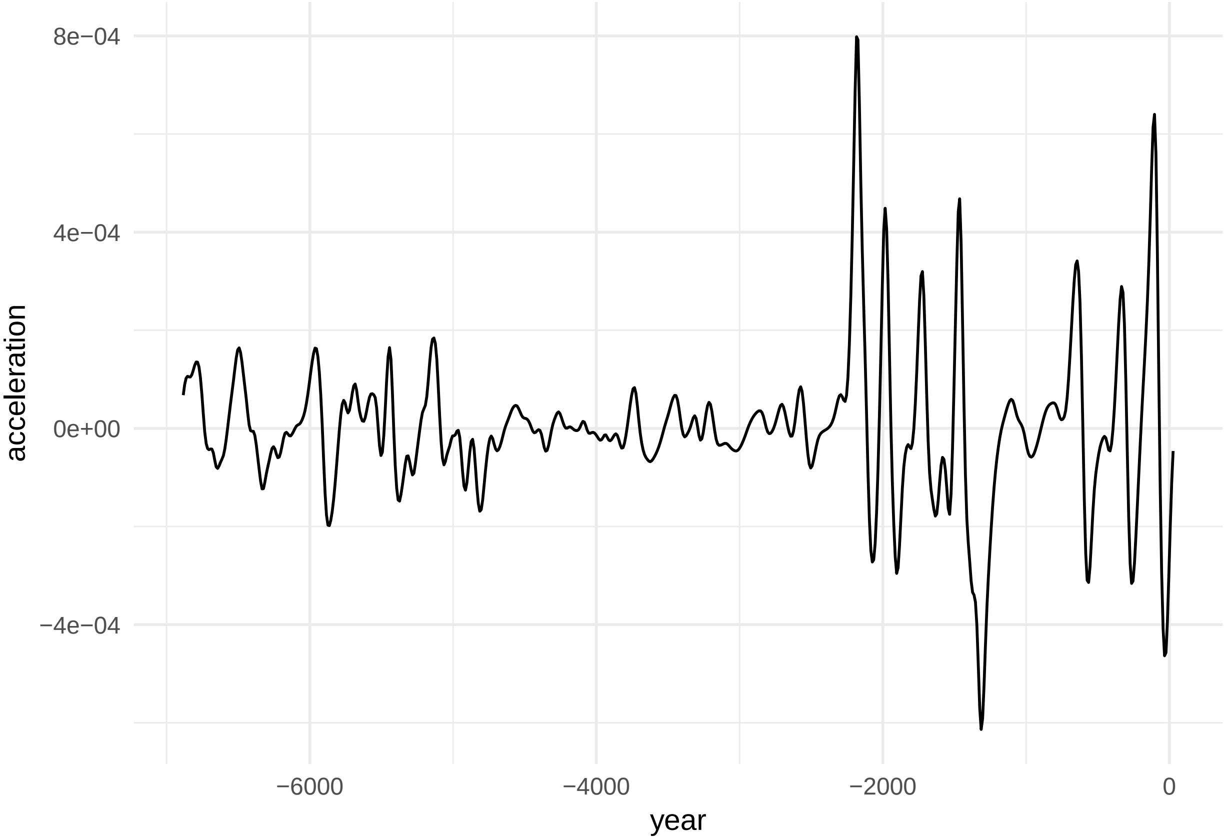
```
## Warning: Removed 7 rows containing missing values (geom_path).
```


150
100
50
0
distance
−6000
−4000
−2000
0
year

1.3.3 Plot the velocity

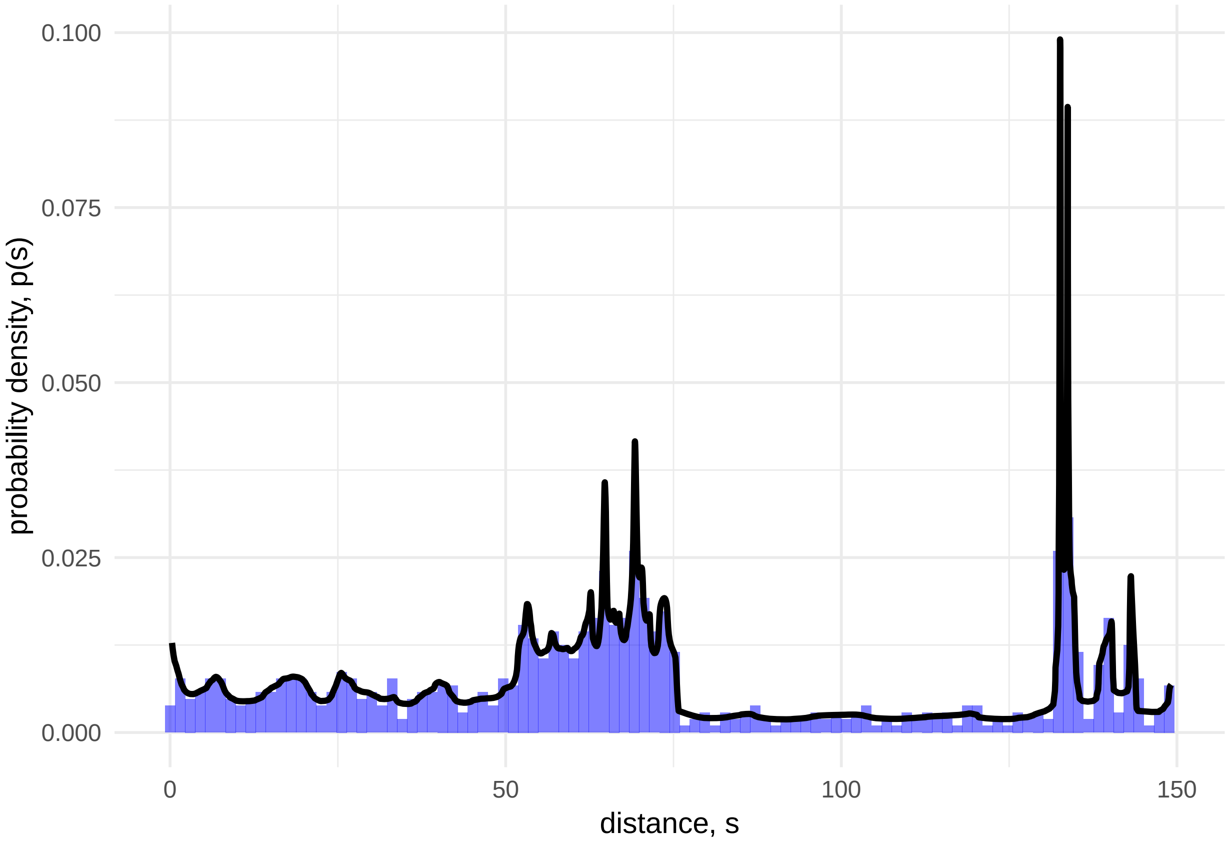


1.3.4 Plot the acceleration



1.3.5 Plot a histogram of distance traveled

Compare histogram of distance travelled to pdf calculated from velocity.



1.4 Moving window analysis

1.4.1 Specifcy parameters for the moving window

Distance over which to move the window (in units time)

```
# Distance over which to move the window (in units time)
winspace <- 50

# Size of the window (in units time)
winsize <- 500

# Start and stop points for windows
t <- distance$YB1950
winStart <- seq(min(t), max(t), by = winspace)
winStop <- winStart + winsize

# Number of windows
nWin <- length(winStart)
```

1.4.2 Loop over data calculating a FI value for each window

```
FI_7.12 <- numeric(length(nWin))
for(i in 1:nWin){
  df <-
    numDiff %>%
    filter(t > winStart[i],
           t <= winStop[i]) %>%
    mutate(TT = max(t) - min(t),
         p = (1/TT)*(1/dsdt))

  FI_7.12[i] <- (1/df$TT[1])*trapz(df$t, df$d2sdt2^2 / df$dsdt^4)
}
```

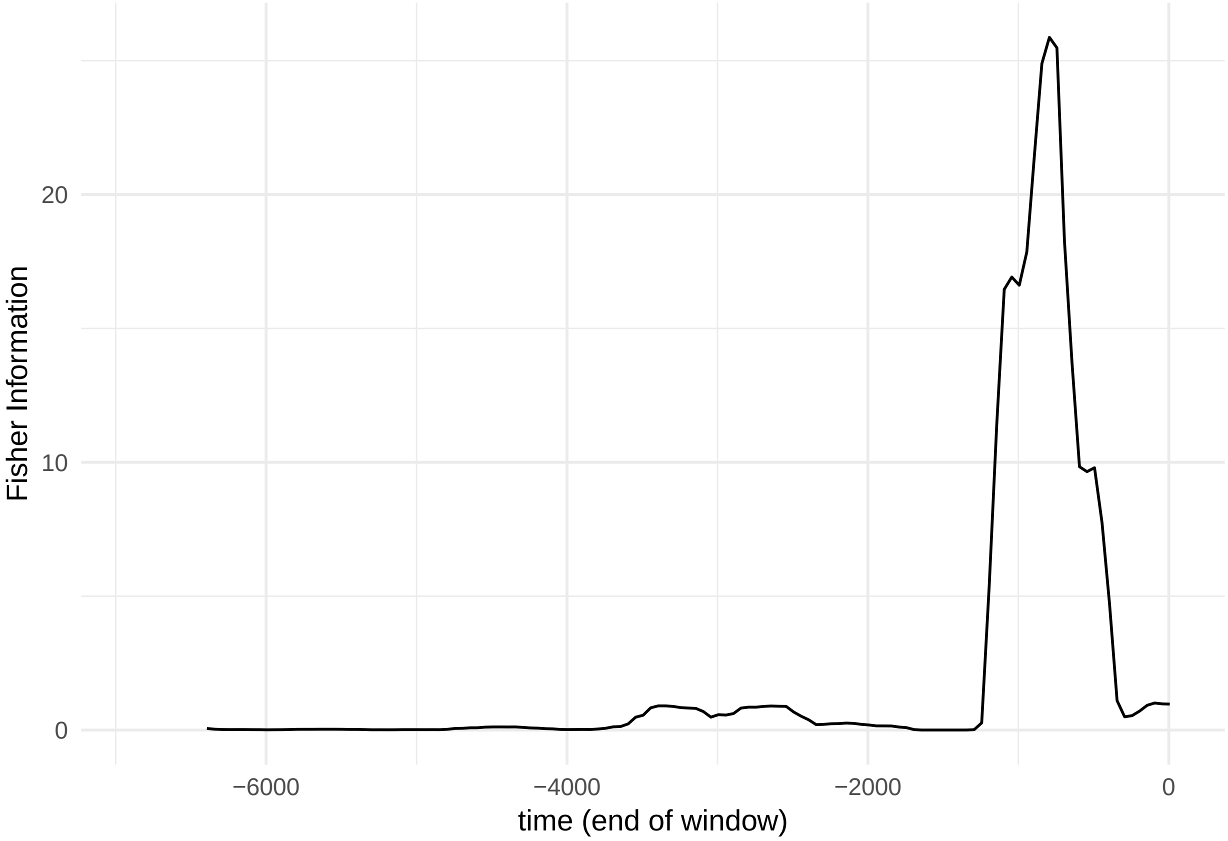
1.5 Plots

1.5.1 Plot Fisher Information for each window

```
# Plot of FI
(gFI <- ggplot(data = data_frame(winStop, FI_7.12),
       aes(x = winStop, y = FI_7.12)) +
  geom_line() +
  labs(x = "time (end of window)",
       y = "Fisher Information") +
  xlim(min(numDiff$t), max(numDiff$t)))
```

```
## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.

## Warning: Removed 10 rows containing missing values (geom_path).
```



```r
# Stacked plot
gB <- ggplotGrob(gS)
```

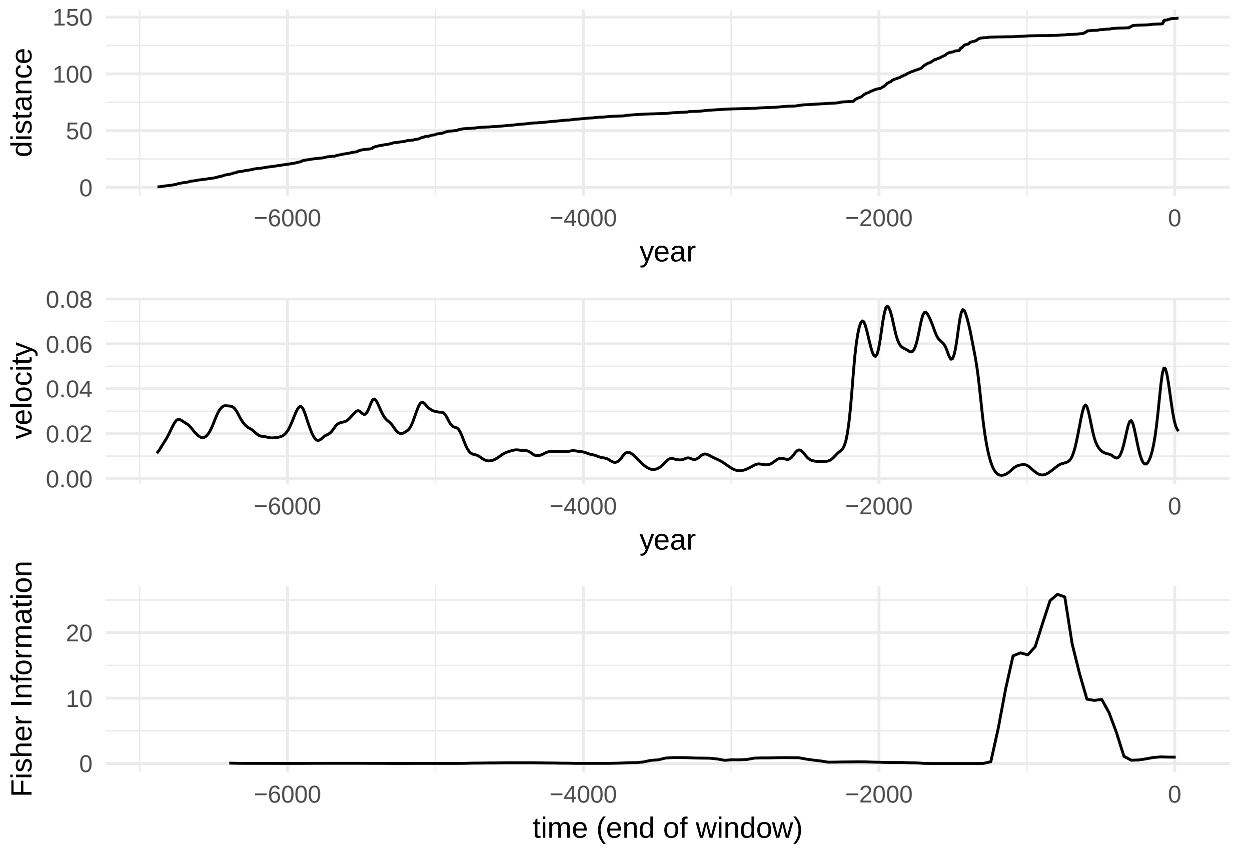
```
## Warning: Removed 7 rows containing missing values (geom_path).
```

```r
gC <- ggplotGrob(gV)
gD <- ggplotGrob(gFI)
```

```
## Warning: Removed 10 rows containing missing values (geom_path).
```

```r
maxWidth = grid::unit.pmax(gB$widths[2:5], gC$widths[2:5], gAcc$widths[2:5], gD$widths[2:5])
gB$widths[2:5] <- as.list(maxWidth)
gC$widths[2:5] <- as.list(maxWidth)
gD$widths[2:5] <- as.list(maxWidth)

grid.arrange(gB, gC, gD, ncol=1)
```



1.5.2 Export figures

```r
# ggsave(
#     "figures/stackplot.png",
#     # plot = grid.arrange(gA, gB, gC, gD, ncol=1) # with relative abundance plot
#     plot = grid.arrange(gB, gC, gD, ncol = 1),
#     units = "cm",
#     dpi = 300
# )
# ggsave(
#     "figures/joyPlotFisher.png",
#     plot = joy,
#     units = "cm",
#     dpi = 300
# )
```

Chartrand, Rick. 2011. “Numerical Differentiation of Noisy, Nonsmooth Data.” Research article. *International Scholarly Research Notices*. https://www.hindawi.com/journals/isrn/2011/164564/.

Spanbauer, Trisha L., Craig R. Allen, David G. Angeler, Tarsha Eason, Sherilyn C. Fritz, Ahjond S. Garmestani, Kirsty L. Nash, and Jeffery R. Stone. 2014. “Prolonged Instability Prior to a Regime Shift.” *PLOS ONE* 9 (10): e108936. doi:10.1371/journal.pone.0108936.