

# APEX active layer analysis

Manies

2022-03-23

Needed libraries:

```
library(lubridate)
library(ggplot2)
library(dplyr)    #left_join
```

## Importation of data files

The data included in this analysis are from various sources. All data are originally from Excel files where a tab was saved in csv format.

```
guelph <- read.csv("raw_data/Seasonal Thaw Depth Data Dielman.csv",header=TRUE)
usgs1 <- read.csv("raw_data/FINAL End of Season usgs.csv",header=TRUE)
crell <- read.csv("raw_data/douglas APEX ERT lines thaw depths October 2017.csv", header=TRUE)
geo.1819.data <- read.csv("raw_data/APEX_2018_2019_Adjusted_ThawDepth_Data.csv", header=TRUE)
geo.now.data <- read.csv("raw_data/APEX_2020_present_ThawDepth_Data.csv", header=TRUE)
apex.data <- read.csv("raw_data/APEX_non_geophysical_thaw_depths.csv", header=TRUE)
region <- read.csv("APEX_GPS_locations.csv", header=TRUE)
```

Add a column to each dataset so we know where the data is from.

```
usgs1 <- cbind(usgs1, source="USGS")
guelph <- cbind(guelph, source="Guelph")
crell <- cbind(crell, source="CRELL")
geo.1819.data <- cbind(geo.1819.data, source="USGS")
geo.now.data <- cbind(geo.now.data, source="USGS")
apex.data <- cbind(apex.data, source="USGS")
```

## Remove bog data

GUELPH:

Names changed throughout the study time. Here old names are converted to current names. Measurements with permafrost are called "Permafrost Plateau" under Site - subset data so only those are chosen.

QUESTION: There are a few sites I assumed were at met\_station region. Confirm with Catherine.

USGS:

Measurements with permafrost are labeled "forest" under Treatment

CRELL:

Type of site was assigned by Kristen Manies based on the descriptions. Anything that said it had live spruce was assumed to have permafrost. Anything labeled thermokarst, noted to have standing water, or dead spruce was assumed to be thawed. For the few rows without descriptions the determination was based on adjacent rows and/or the depth measurements.

Once imported each file needed their time stamp converted into strptime so can combine them later.

```
lookup <- read.csv("raw_data/guelph_lookup.csv",header=TRUE)
lookup$Plot <- as.character(lookup$Plot)
guelph <- left_join(guelph, lookup, by = c("Treatment" , "Plot"))
guelph$new_plot <- ifelse(is.na(guelph$new_plot),guelph$Plot,guelph$new_plot)

guelph.gamma <- subset(guelph, Site=="Permafrost Plateau ")

crell.gamma <- subset(crell, Site=="Pfrost")
crell.gamma$date <- substr(crell.gamma$DateTime, 1, 10)
crell.gamma$transect <- substr(crell.gamma$Name, 4,5)
crell.gamma$year <- year(crell.gamma$date)

usgs.gamma <- subset(usgs1, usgs1$Treatment == "forest")
geo.1819.gamma <- subset(geo.1819.data, SiteID != "APEX4" | SiteID != "APEX5")
geo.now.gamma <- subset(geo.now.data, SiteID != "APEX4" | SiteID != "APEX5")

guelph.gamma$Date <- strptime(guelph.gamma$Date, format="%m/%d/%Y", tz="America/Anchorage")
usgs.gamma$date <- strptime(usgs.gamma$date, format="%d-%b-%y", tz="America/Anchorage")
crell.gamma$date <- substr(crell.gamma$DateTime, 1, 10)
crell.gamma$date <- strptime(crell.gamma$date, format="%Y-%m-%d")
geo.1819.gamma$Date <- paste(substr(geo.1819.gamma$Date, 1, 4), substr(geo.1819.gamma$Date, 5, 6
), substr(geo.1819.gamma$Date, 7, 8), sep = '-')
geo.1819.gamma$Date <- strptime(geo.1819.gamma$Date, format="%Y-%m-%d", tz="America/Anchorage")
geo.now.gamma$Date <- strptime(geo.now.gamma$Date, format="%m/%d/%Y", tz="America/Anchorage")
apex.data$Date <- strptime(apex.data$Date, format="%m/%d/%Y", tz="America/Anchorage")
```

## Combining all datasets

Combine gamma data from all three data sets and plot over time. Subset larger dataframe to just columns we need. Columns are: site (aka within gamma location), location (more detailed location), Date (%Y-%m-%d format), active layer depth, & source. Using a function, give all dataframes the same names so they can be combined.

```

cg.sub <- data.frame(crell.gamma$transect, crell.gamma$GPSpoint, crell.gamma$date, crell.gamma$Depth, crell.gamma$source)
gg.sub <- data.frame(guelph.gamma$Treatment, guelph.gamma$new_plot, guelph.gamma$Date, guelph.gamma$Ice_Depth_cm, guelph.gamma$source)
usgs.sub <- data.frame(usgs.gamma$Site, usgs.gamma$Chamber, usgs.gamma$date, usgs.gamma$AL_cm, usgs.gamma$source)
geo.1819.sub <- data.frame(geo.1819.gamma$SiteID, geo.1819.gamma$Adjacency, geo.1819.gamma$Date, geo.1819.gamma$ThawDep_cm, geo.1819.gamma$source)
geo.now.sub <- data.frame(geo.now.gamma$SiteID, geo.now.gamma$Adjacency, geo.now.gamma$Date, geo.now.data$ThawDep_cm, geo.now.gamma$source)
apex.sub <- data.frame(apex.data$Site, apex.data$Location, apex.data$Date, apex.data$AL_cm, apex.data$source)

same.names <- function(df) {
  names(df)[1] <- "site"
  names(df)[2] <- "location"
  names(df)[3] <- "date"
  names(df)[4] <- "depth"
  names(df)[5] <- "source"
  return(df)
}

cg.sub <- same.names(cg.sub)
gg.sub <- same.names(gg.sub)
usgs.sub <- same.names(usgs.sub)
geo.1819.sub <- same.names(geo.1819.sub)
geo.now.sub <- same.names(geo.now.sub)
apex.sub <- same.names(apex.sub)
all.gamma <- rbind(cg.sub, gg.sub, usgs.sub, geo.1819.sub, geo.now.sub, apex.sub)

```

## Assign each site-location a region within APEX

So we can look at data by location within APEX to see if there are trends. Choices are: beta\_tower\_forest, emergent\_bog\_area, met\_station, NE\_of\_emergent\_bog (BZFM area), NW\_of\_emergent\_bog (BZFP area), tower\_to\_old\_taj, and towards\_alpha. Remove data from the BZBB area (emergent\_bog\_area) and from the CRELL data where there are only one or two measurements for that location for that year (too little data to feel secure in) or there is only that year of data (towards\_alpha)

```

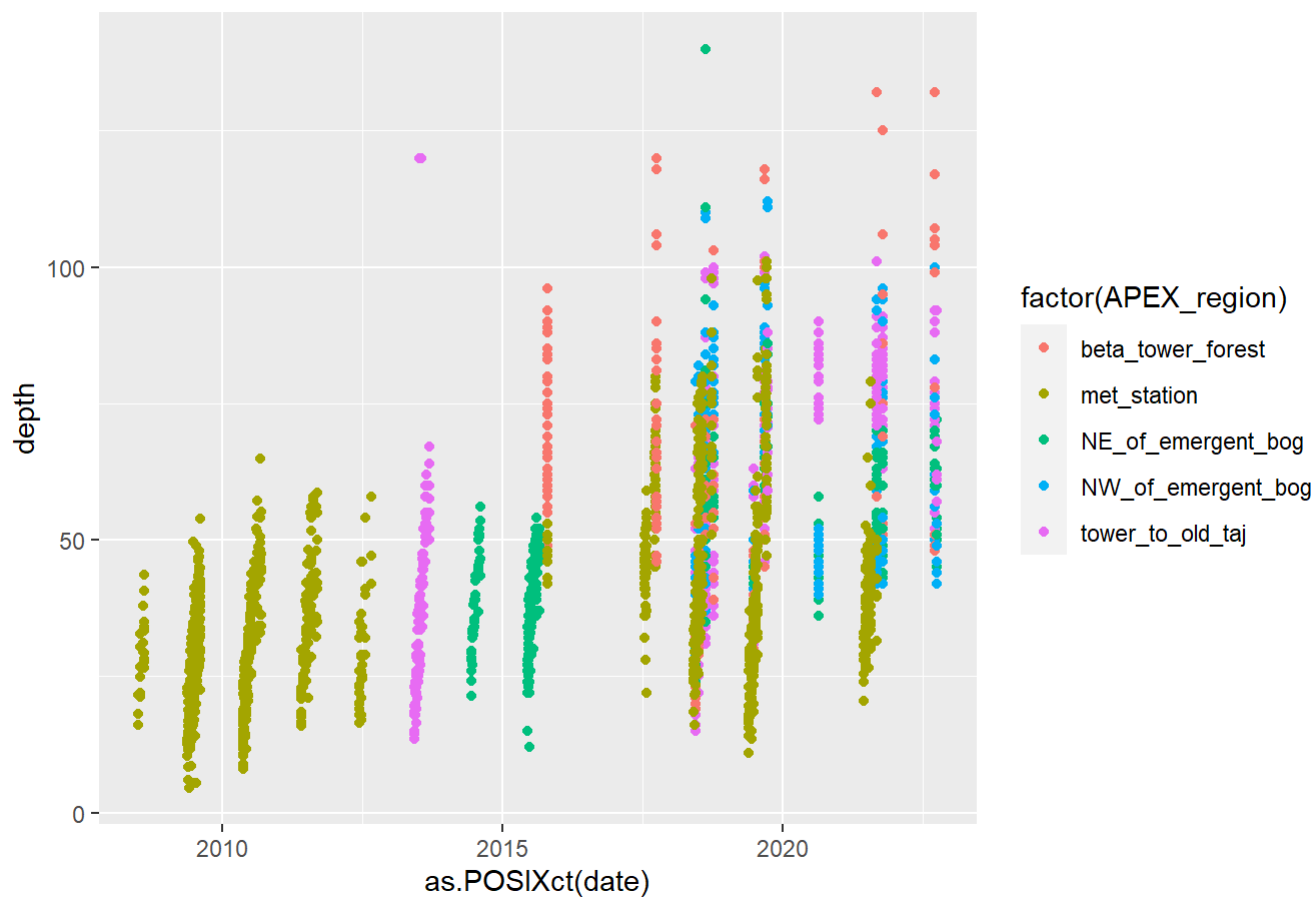
all.region <- merge(all.gamma, region, by=c("site", "location"))
all.nonbog <- subset(all.region, APEX_region != "emergent_bog_area")
all.nonbog <- subset(all.nonbog, location!="439")
all.nonbog <- subset(all.nonbog, location!="440")
all.nonbog <- subset(all.nonbog, location!="441")
all.nonbog <- subset(all.nonbog, APEX_region != "towards_alpha")

all.plot <- ggplot() +
  geom_point(data=all.nonbog, aes(x=as.POSIXct(date), y=depth, color=factor(APEX_region))) +
  ggtitle("All dates")
print(all.plot)

```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```

All dates



## Active layer data

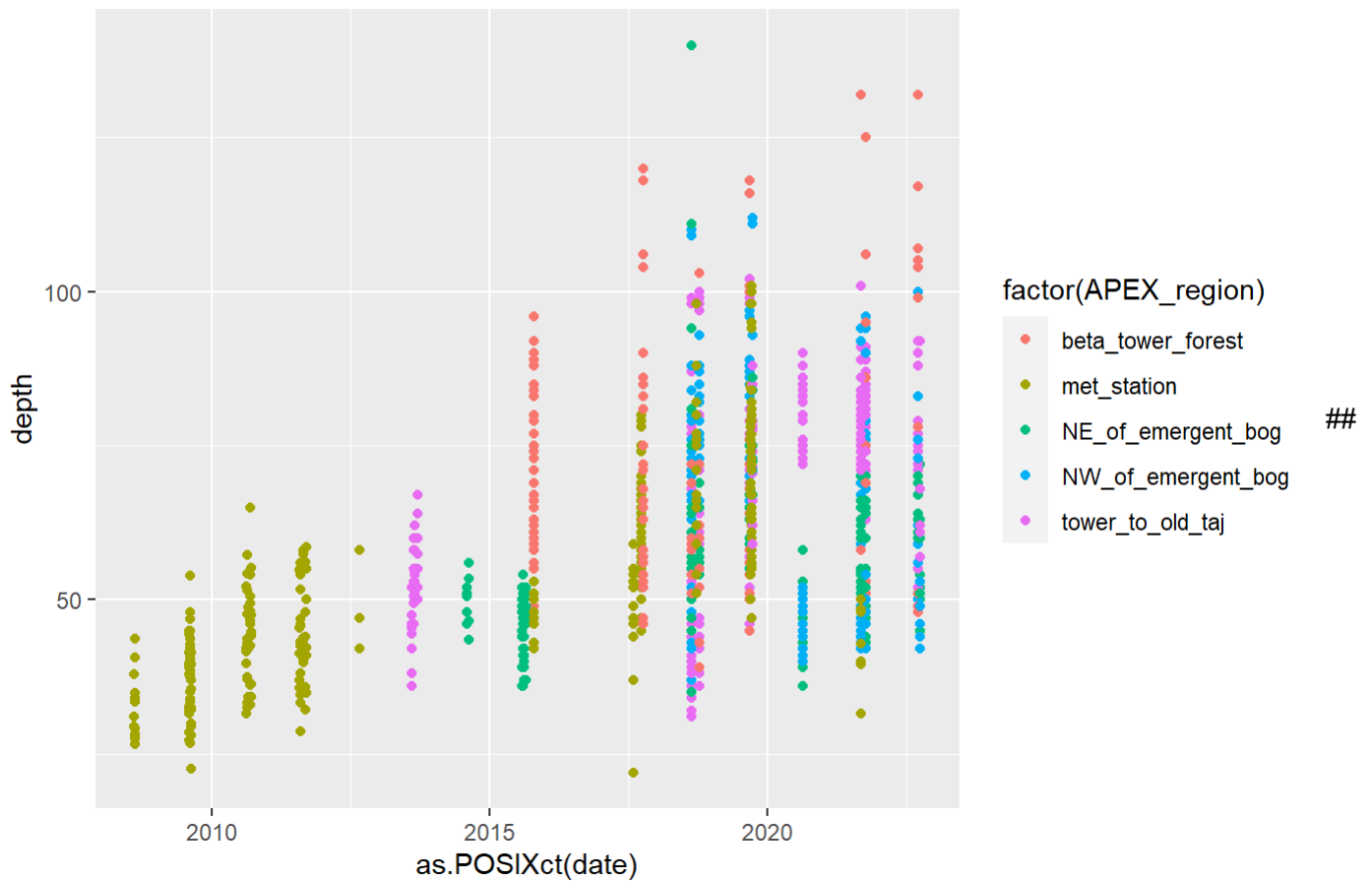
Only looking at data from August through October, so data not impacted by seasonal ice.

```
august.on <- subset(all.nonbog, month(date)>=8)

aug.plot <- ggplot(data=august.on, aes(x=as.POSIXct(date), y=depth)) +
  geom_point(aes(color=factor(APEX_region))) +
  ggtitle("Active layer (Aug - Oct)")
print(aug.plot)
```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```

## Active layer (Aug - Oct)



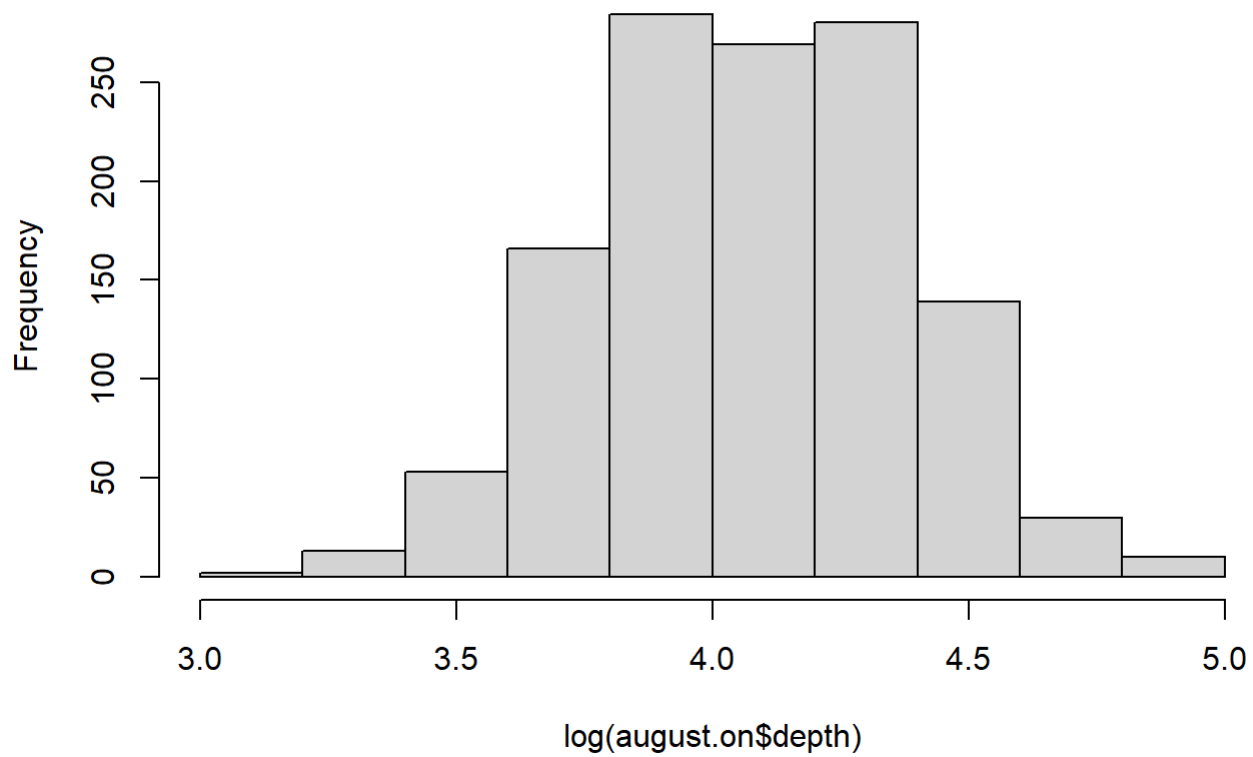
Statistics on the data

ADD POST-HOC TEST?

```
august.on$year <- year(as.POSIXct(august.on$date, format="%Y-%m-%d", tz="America/Anchorage"))

#What's the distribution of the data?
#Log helps normality
hist(log(august.on$depth))
```

**Histogram of log(august.on\$depth)**



```
AL.region.stats <- glm(data=august.on, log(depth) ~ factor(year) * APEX_region)
summary(AL.region.stats)
```

```
##
## Call:
## glm(formula = log(depth) ~ factor(year) * APEX_region, data = august.on)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95575  -0.13581   0.00986   0.12907   0.86736
##
## Coefficients: (36 not defined because of singularities)
##                                Estimate Std. Error t value
## (Intercept)                   4.12536    0.10400  39.666
## factor(year)2009               0.12252    0.05898   2.077
## factor(year)2010               0.27255    0.06073   4.488
## factor(year)2011               0.25819    0.06148   4.200
## factor(year)2012               0.38199    0.13165   2.902
## factor(year)2013              -0.08101    0.13139  -0.617
## factor(year)2014               0.09614    0.14334   0.671
## factor(year)2015               0.08045    0.10770   0.747
## factor(year)2017               0.11780    0.10879   1.083
## factor(year)2018              -0.03629    0.11211  -0.324
## factor(year)2019               0.10428    0.11720   0.890
## factor(year)2020               0.38267    0.13478   2.839
## factor(year)2021               0.24764    0.09483   2.612
## factor(year)2022               0.27045    0.12027   2.249
## APEX_regionmet_station         -0.62460    0.08989  -6.949
## APEX_regionNE_of_emergent_bog  -0.33040    0.07798  -4.237
## APEX_regionNW_of_emergent_bog  -0.31327    0.07890  -3.971
## APEX_regiontower_to_old_taj    -0.11331    0.07183  -1.578
## factor(year)2009:APEX_regionmet_station      NA         NA      NA
## factor(year)2010:APEX_regionmet_station      NA         NA      NA
## factor(year)2011:APEX_regionmet_station      NA         NA      NA
## factor(year)2012:APEX_regionmet_station      NA         NA      NA
## factor(year)2013:APEX_regionmet_station      NA         NA      NA
## factor(year)2014:APEX_regionmet_station      NA         NA      NA
## factor(year)2015:APEX_regionmet_station      0.27006    0.11507   2.347
## factor(year)2017:APEX_regionmet_station      0.42823    0.09902   4.325
## factor(year)2018:APEX_regionmet_station      0.79452    0.11211   7.087
## factor(year)2019:APEX_regionmet_station      0.64778    0.10649   6.083
## factor(year)2020:APEX_regionmet_station      NA         NA      NA
## factor(year)2021:APEX_regionmet_station      NA         NA      NA
## factor(year)2022:APEX_regionmet_station      NA         NA      NA
## factor(year)2009:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2010:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2011:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2012:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2013:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2014:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2015:APEX_regionNE_of_emergent_bog -0.05181    0.08778  -0.590
## factor(year)2017:APEX_regionNE_of_emergent_bog NA         NA      NA
## factor(year)2018:APEX_regionNE_of_emergent_bog 0.31561    0.09615   3.282
## factor(year)2019:APEX_regionNE_of_emergent_bog 0.30838    0.10918   2.825
## factor(year)2020:APEX_regionNE_of_emergent_bog -0.36136    0.09122  -3.961
```

## factor(year)2021:APEX_regionNE_of_emergent_bog	-0.05176	0.09353	-0.553
## factor(year)2022:APEX_regionNE_of_emergent_bog	NA	NA	NA
## factor(year)2009:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2010:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2011:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2012:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2013:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2014:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2015:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2017:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2018:APEX_regionNW_of_emergent_bog	0.41161	0.09306	4.423
## factor(year)2019:APEX_regionNW_of_emergent_bog	0.50505	0.10438	4.839
## factor(year)2020:APEX_regionNW_of_emergent_bog	-0.36753	0.09200	-3.995
## factor(year)2021:APEX_regionNW_of_emergent_bog	-0.01190	0.09351	-0.127
## factor(year)2022:APEX_regionNW_of_emergent_bog	NA	NA	NA
## factor(year)2009:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2010:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2011:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2012:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2013:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2014:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2015:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2017:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2018:APEX_regiontower_to_old_taj	0.06186	0.08588	0.720
## factor(year)2019:APEX_regiontower_to_old_taj	0.13749	0.09514	1.445
## factor(year)2020:APEX_regiontower_to_old_taj	NA	NA	NA
## factor(year)2021:APEX_regiontower_to_old_taj	0.08809	0.08718	1.010
## factor(year)2022:APEX_regiontower_to_old_taj	NA	NA	NA
##	Pr(> t )		
## (Intercept)	< 2e-16	***	
## factor(year)2009	0.03799	*	
## factor(year)2010	7.89e-06	***	
## factor(year)2011	2.87e-05	***	
## factor(year)2012	0.00378	**	
## factor(year)2013	0.53765		
## factor(year)2014	0.50254		
## factor(year)2015	0.45520		
## factor(year)2017	0.27911		
## factor(year)2018	0.74619		
## factor(year)2019	0.37377		
## factor(year)2020	0.00460	**	
## factor(year)2021	0.00913	**	
## factor(year)2022	0.02471	*	
## APEX_regionmet_station	6.01e-12	***	
## APEX_regionNE_of_emergent_bog	2.44e-05	***	
## APEX_regionNW_of_emergent_bog	7.59e-05	***	
## APEX_regiontower_to_old_taj	0.11493		
## factor(year)2009:APEX_regionmet_station	NA		
## factor(year)2010:APEX_regionmet_station	NA		
## factor(year)2011:APEX_regionmet_station	NA		
## factor(year)2012:APEX_regionmet_station	NA		
## factor(year)2013:APEX_regionmet_station	NA		



```

## factor(year)2014:APEX_regionmet_station      NA
## factor(year)2015:APEX_regionmet_station      0.01909 *
## factor(year)2017:APEX_regionmet_station      1.65e-05 ***
## factor(year)2018:APEX_regionmet_station      2.32e-12 ***
## factor(year)2019:APEX_regionmet_station      1.58e-09 ***
## factor(year)2020:APEX_regionmet_station      NA
## factor(year)2021:APEX_regionmet_station      NA
## factor(year)2022:APEX_regionmet_station      NA
## factor(year)2009:APEX_regionNE_of_emergent_bog NA
## factor(year)2010:APEX_regionNE_of_emergent_bog NA
## factor(year)2011:APEX_regionNE_of_emergent_bog NA
## factor(year)2012:APEX_regionNE_of_emergent_bog NA
## factor(year)2013:APEX_regionNE_of_emergent_bog NA
## factor(year)2014:APEX_regionNE_of_emergent_bog NA
## factor(year)2015:APEX_regionNE_of_emergent_bog 0.55512
## factor(year)2017:APEX_regionNE_of_emergent_bog NA
## factor(year)2018:APEX_regionNE_of_emergent_bog 0.00106 **
## factor(year)2019:APEX_regionNE_of_emergent_bog 0.00481 **
## factor(year)2020:APEX_regionNE_of_emergent_bog 7.89e-05 ***
## factor(year)2021:APEX_regionNE_of_emergent_bog 0.58012
## factor(year)2022:APEX_regionNE_of_emergent_bog NA
## factor(year)2009:APEX_regionNW_of_emergent_bog NA
## factor(year)2010:APEX_regionNW_of_emergent_bog NA
## factor(year)2011:APEX_regionNW_of_emergent_bog NA
## factor(year)2012:APEX_regionNW_of_emergent_bog NA
## factor(year)2013:APEX_regionNW_of_emergent_bog NA
## factor(year)2014:APEX_regionNW_of_emergent_bog NA
## factor(year)2015:APEX_regionNW_of_emergent_bog NA
## factor(year)2017:APEX_regionNW_of_emergent_bog NA
## factor(year)2018:APEX_regionNW_of_emergent_bog 1.06e-05 ***
## factor(year)2019:APEX_regionNW_of_emergent_bog 1.48e-06 ***
## factor(year)2020:APEX_regionNW_of_emergent_bog 6.87e-05 ***
## factor(year)2021:APEX_regionNW_of_emergent_bog 0.89876
## factor(year)2022:APEX_regionNW_of_emergent_bog NA
## factor(year)2009:APEX_regiontower_to_old_taj  NA
## factor(year)2010:APEX_regiontower_to_old_taj  NA
## factor(year)2011:APEX_regiontower_to_old_taj  NA
## factor(year)2012:APEX_regiontower_to_old_taj  NA
## factor(year)2013:APEX_regiontower_to_old_taj  NA
## factor(year)2014:APEX_regiontower_to_old_taj  NA
## factor(year)2015:APEX_regiontower_to_old_taj  NA
## factor(year)2017:APEX_regiontower_to_old_taj  NA
## factor(year)2018:APEX_regiontower_to_old_taj  0.47147
## factor(year)2019:APEX_regiontower_to_old_taj  0.14866
## factor(year)2020:APEX_regiontower_to_old_taj  NA
## factor(year)2021:APEX_regiontower_to_old_taj  0.31246
## factor(year)2022:APEX_regiontower_to_old_taj  NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04378771)
##

```

```
## Null deviance: 112.266 on 1245 degrees of freedom
## Residual deviance: 53.071 on 1212 degrees of freedom
## (3 observations deleted due to missingness)
## AIC: -326.47
##
## Number of Fisher Scoring iterations: 2
```

```
source("summarystats.R")
#For all regions and years
year.region.stats <- summarystats(august.on, measurevar="depth", groupvars=c("APEX_region", "year"), na.rm=TRUE)
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## -----
```

```
##
## Attaching package: 'plyr'
```

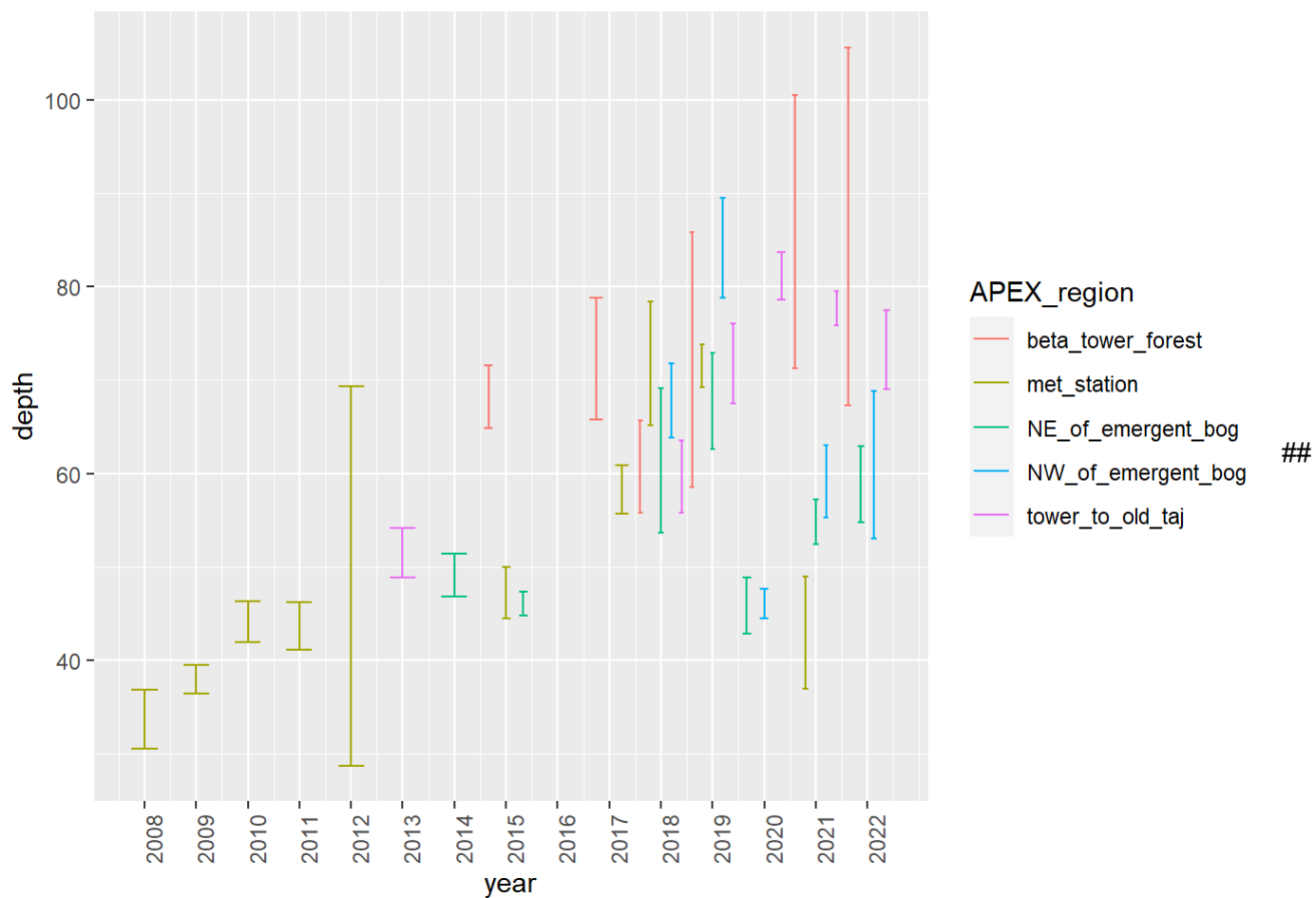
```
## The following objects are masked from 'package:dplyr':
##
## arrange, count, desc, failwith, id, mutate, rename, summarise,
## summarize
```

```
print(year.region.stats)
```

##	APEX_region	year	N	depth	sd	se	ci
## 1	beta_tower_forest	2015	56	68.19643	12.674125	1.6936512	3.394153
## 2	beta_tower_forest	2017	43	72.32558	21.211768	3.2347643	6.528019
## 3	beta_tower_forest	2018	25	60.72000	12.077941	2.4155883	4.985529
## 4	beta_tower_forest	2019	15	72.20000	24.678504	6.3719624	13.666500
## 5	beta_tower_forest	2021	24	85.87500	34.689069	7.0808766	14.647909
## 6	beta_tower_forest	2022	12	86.50000	30.164699	8.7077987	19.165736
## 7	met_station	2008	16	33.61250	5.941478	1.4853696	3.165990
## 8	met_station	2009	59	37.91186	5.729990	0.7459812	1.493244
## 9	met_station	2010	46	44.12609	7.389495	1.0895216	2.194409
## 10	met_station	2011	42	43.62381	8.047112	1.2416963	2.507656
## 11	met_station	2012	3	49.00000	8.185353	4.7258156	20.333544
## 12	met_station	2015	10	47.20000	3.910101	1.2364825	2.797118
## 13	met_station	2017	62	58.26613	10.431556	1.3248089	2.649119
## 14	met_station	2018	16	71.75000	12.423097	3.1057742	6.619801
## 15	met_station	2019	128	71.48438	13.152338	1.1625135	2.300404
## 16	met_station	2021	7	42.91429	6.553734	2.4770786	6.061193
## 17	NE_of_emergent_bog	2014	12	49.08333	3.591868	1.0368830	2.282164
## 18	NE_of_emergent_bog	2015	52	46.00000	4.537426	0.6292278	1.263227
## 19	NE_of_emergent_bog	2018	31	61.38710	21.024870	3.7761781	7.711985
## 20	NE_of_emergent_bog	2019	15	67.76667	9.263729	2.3918844	5.130082
## 21	NE_of_emergent_bog	2020	20	45.85000	6.425893	1.4368734	3.007411
## 22	NE_of_emergent_bog	2021	52	54.75962	8.606023	1.1934407	2.395932
## 23	NE_of_emergent_bog	2022	18	58.83333	8.161963	1.9237932	4.058849
## 24	NW_of_emergent_bog	2018	64	67.78125	15.845954	1.9807442	3.958201
## 25	NW_of_emergent_bog	2019	25	84.20000	13.038405	2.6076810	5.381989
## 26	NW_of_emergent_bog	2020	20	46.05000	3.332061	0.7450715	1.559453
## 27	NW_of_emergent_bog	2021	63	59.09524	15.356741	1.9347675	3.867545
## 28	NW_of_emergent_bog	2022	17	60.94118	15.364206	3.7263673	7.899546
## 29	tower_to_old_taj	2013	34	51.52941	7.596791	1.3028389	2.650646
## 30	tower_to_old_taj	2018	94	59.64894	18.908215	1.9502335	3.872777
## 31	tower_to_old_taj	2019	45	71.76667	14.283096	2.1291982	4.291117
## 32	tower_to_old_taj	2020	20	81.20000	5.463756	1.2217330	2.557117
## 33	tower_to_old_taj	2021	71	77.67606	7.671421	0.9104302	1.815796
## 34	tower_to_old_taj	2022	29	73.27586	11.218278	2.0831819	4.267205

```
write.csv(year.region.stats, "year.region.stats.csv")
```

```
plot.by.year <- ggplot(year.region.stats, aes(x=year, y=depth, color=APEX_region)) +
  geom_errorbar(aes(ymin=depth-ci, ymax=depth+ci), width=0.5, position=position_dodge(width=1))
+
  scale_x_continuous(breaks=seq(2008, 2022, 1)) +
  theme(axis.text.x = element_text(angle = 90))
print(plot.by.year)
```



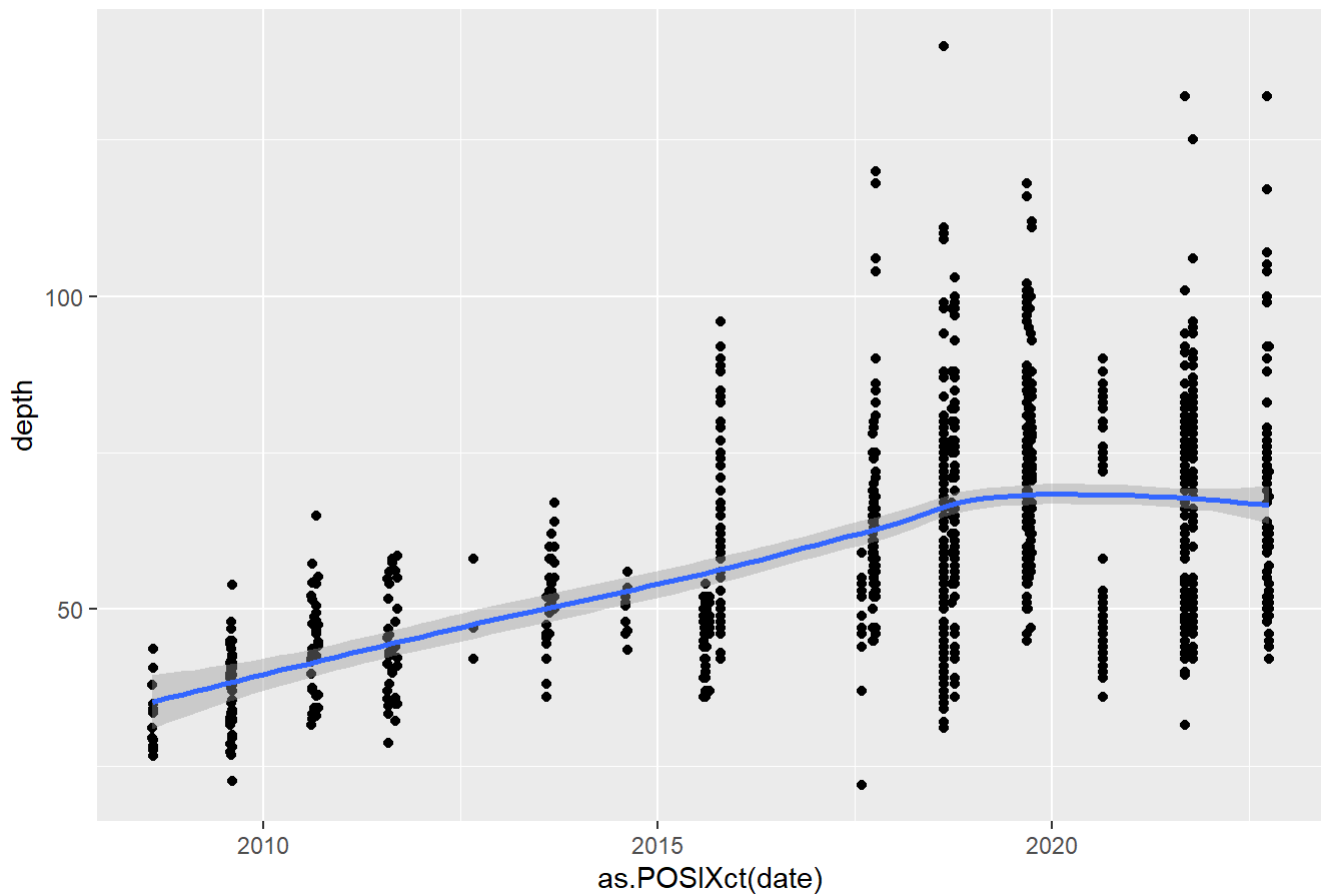
Plot all active lyear data with error

```
error.plot <- ggplot(data=august.on) +
  geom_point(aes(x = as.POSIXct(date), y = depth)) +
  geom_smooth(aes(x = as.POSIXct(date), y = depth), method = "loess", formula = y ~ x) +
  ggtitle("Active layer over time at APEX")
print(error.plot)
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```

## Active layer over time at APEX



```
region.error.plot <- ggplot(data=august.on) +  
  geom_point(aes(x = as.POSIXct(date), y = depth, color=APEX_region)) +  
  geom_smooth(aes(x = as.POSIXct(date), y = depth, color=APEX_region), method = "lm", formula =  
    y ~ x) +  
  ggtitle("Active layer over time at APEX")  
print(region.error.plot)
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).  
## Removed 3 rows containing missing values (`geom_point()`).
```

Active layer over time at APEX

