

# More DSM residence time analyss

Rosie

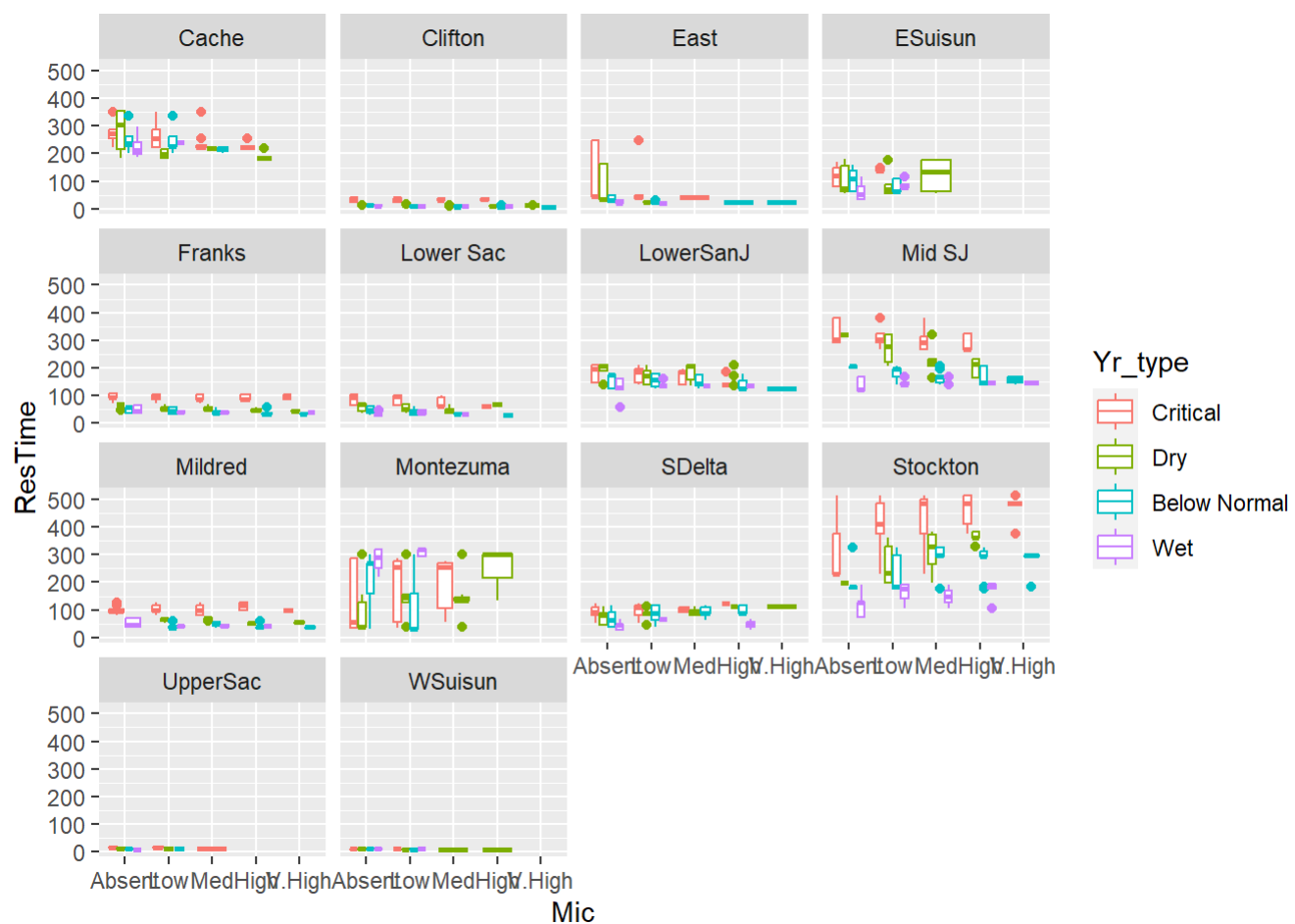
2023-08-08

My models were all a bit of a mess. It's not really working very well with the residence time values. The salinity relationship makes no sense. I'm going to limit the dataset to just the past five years. I may want to limit it to just the delta and kick out SUIsun

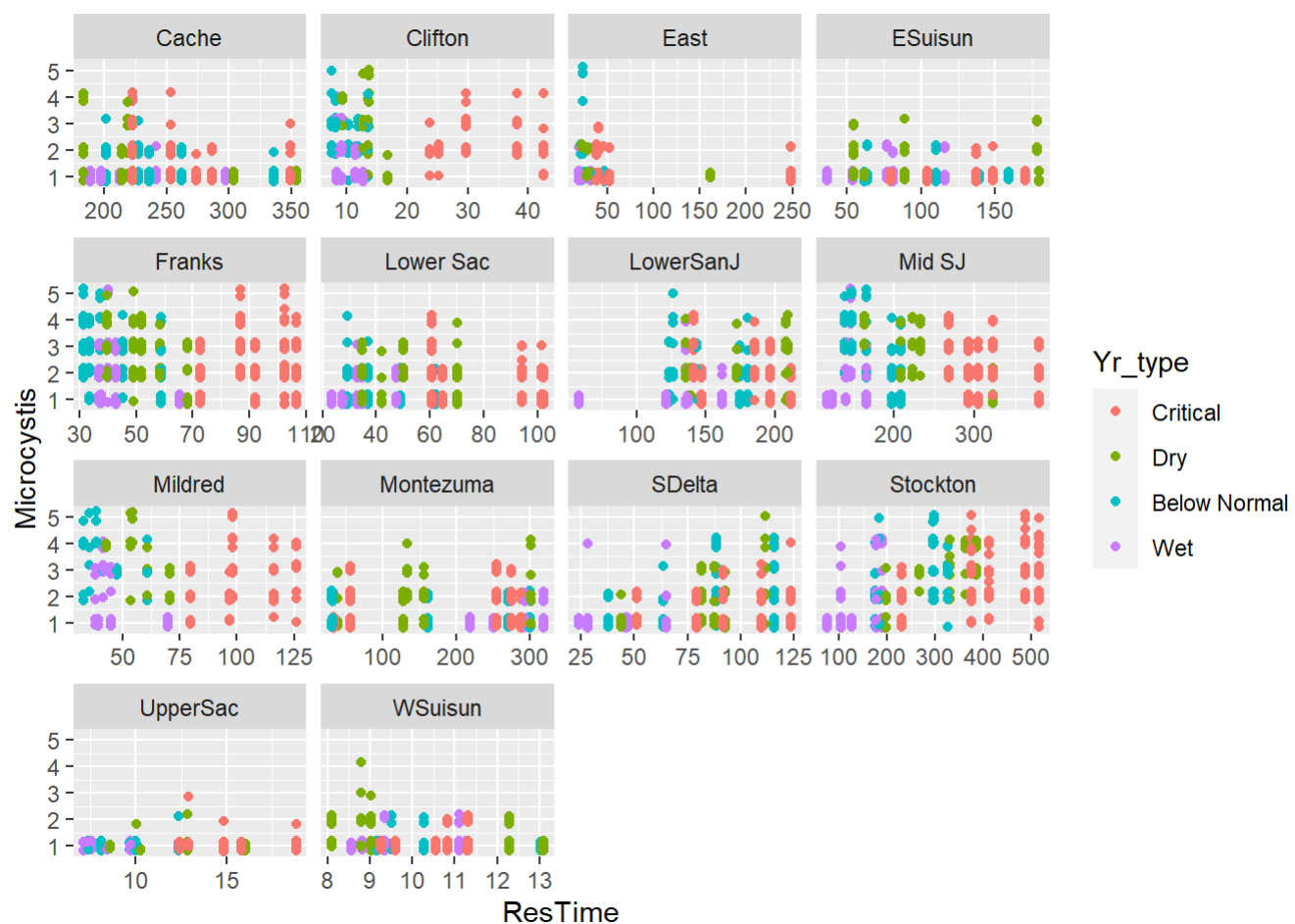
```
load("data/HABrestime.RData")

HABres2 = HABrestime %>%
  st_drop_geometry() %>%
  filter(Year>2014, !is.na(Temperature), !is.na(ResTime), !is.na(Region),
         !is.na(Station), !is.na(Microcystis), Month %in% c(6:10)) %>%
  mutate(MicF = case_when(Microcystis ==1 ~ "Absent",
                          Microcystis %in% c(2, 2.5, 3, 3.5)~ "Mid",
                          Microcystis %in% c(4, 4.5,5) ~ "High"),
         MicF = factor(MicF, levels = c("Absent", "Mid", "High"), ordered = T))

#boxplot of residence time by microcystis scores
ggplot(HABres2, aes(x =Mic, y = ResTime, color = Yr_type)) +
  geom_boxplot()+
  facet_wrap(~Region)
```



```
#scatterplot of microcystis scores by residence time
ggplot(HABres2, aes(x =ResTime, y = Microcystis, color = Yr_type)) +
  geom_point(position = "jitter")+
  facet_wrap(~Region, scales = "free_x")
```



Now an ordered regression

```
m1 = clmm(MicF ~ ResTime+ Temperature+ Month + (1|Station)+ (1|Year), data = HABres2)
```

```
summary(m1)
```

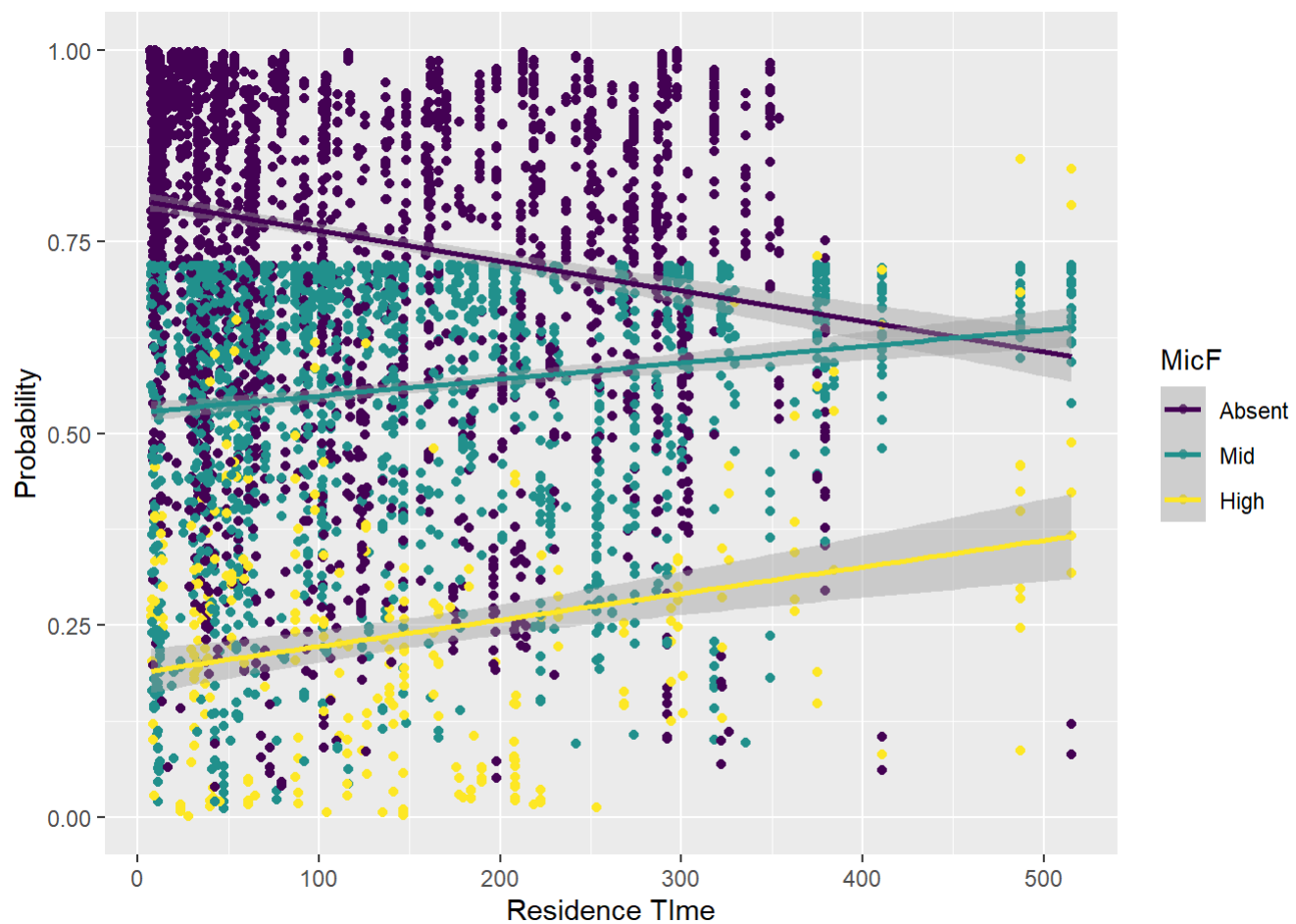
```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: MicF ~ ResTime + Temperature + Month + (1 | Station) + (1 | Year)
## data:    HABres2
##
## link threshold nobs logLik  AIC      niter      max.grad cond.H
## logit flexible  4750 -2902.43 5818.87 1399(6983) 1.12e+02 4.8e+06
##
## Random effects:
## Groups Name          Variance Std.Dev.
## Station (Intercept) 1.946     1.395
## Year (Intercept)    1.051     1.025
## Number of groups:  Station 169, Year 8
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## ResTime      0.0002216  0.0005381   0.412    0.68
## Temperature  0.4599304  0.0232555  19.777 <2e-16 ***
## Month        0.3148033  0.0325501   9.671 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##              Estimate Std. Error z value
## Absent|Mid   12.8466     0.7887  16.29
## Mid|High     16.4707     0.8111  20.31
```

```
#not as exciting as I'd think
#Also, i need to work on how to plot this.
```

```
HABres2$fits = fitted(m1)
```

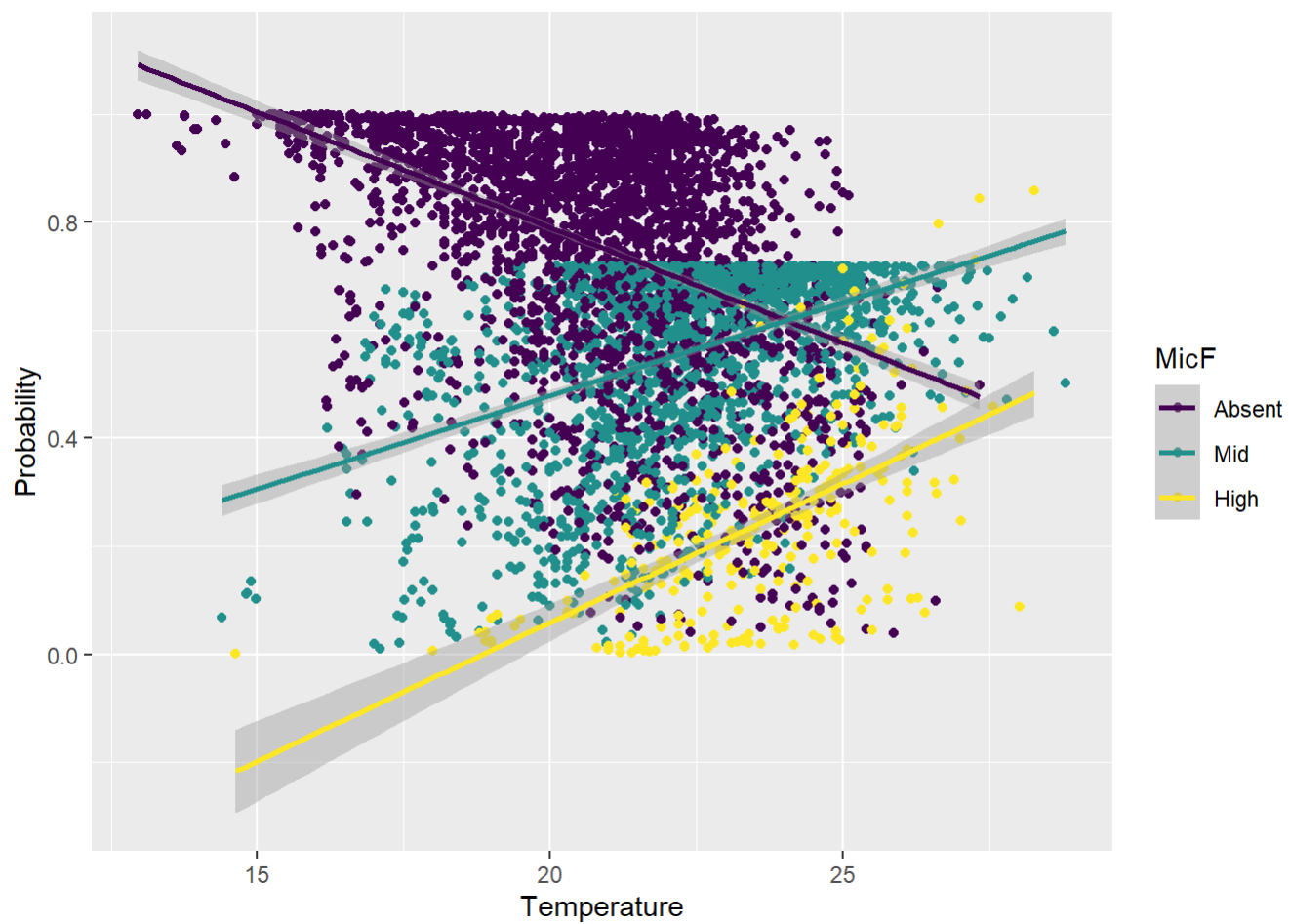
```
ggplot(HABres2, aes( ResTime, fits, color = MicF))+ geom_point()+ geom_smooth(method = "lm")+
  ylab("Probability")+ xlab("Residence TIme")
```

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

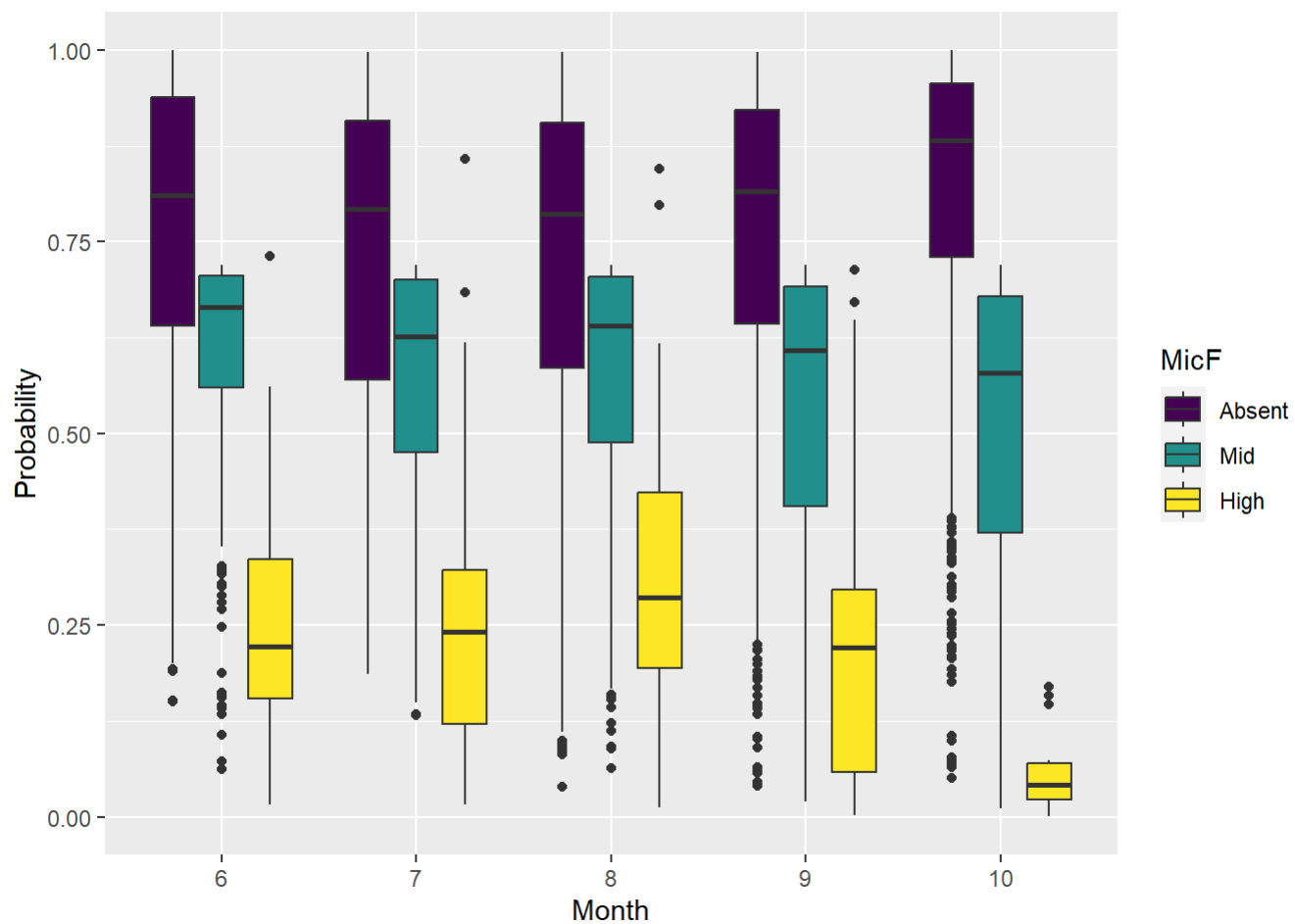


```
ggplot(HABres2, aes(Temperature, fits, color = MicF))+ geom_point()+ geom_smooth(method = "lm")+
  ylab("Probability")+ xlab("Temperature")
```

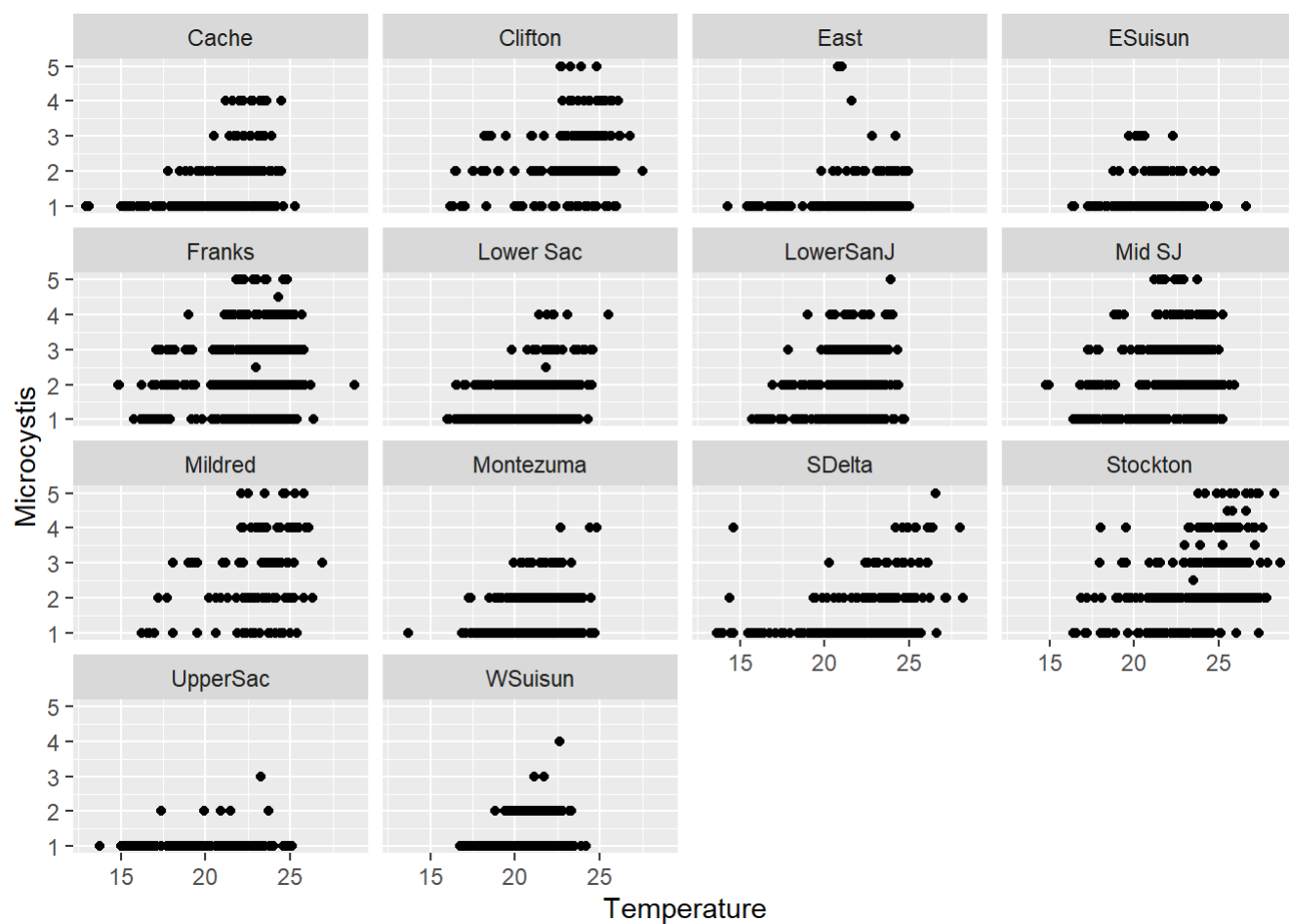
```
## `geom_smooth()` using formula = 'y ~ x'
```



```
ggplot(HABres2, aes(as.factor(Month), fits, fill = MicF))+ geom_boxplot()+  
  ylab("Probability")+ xlab("Month")
```

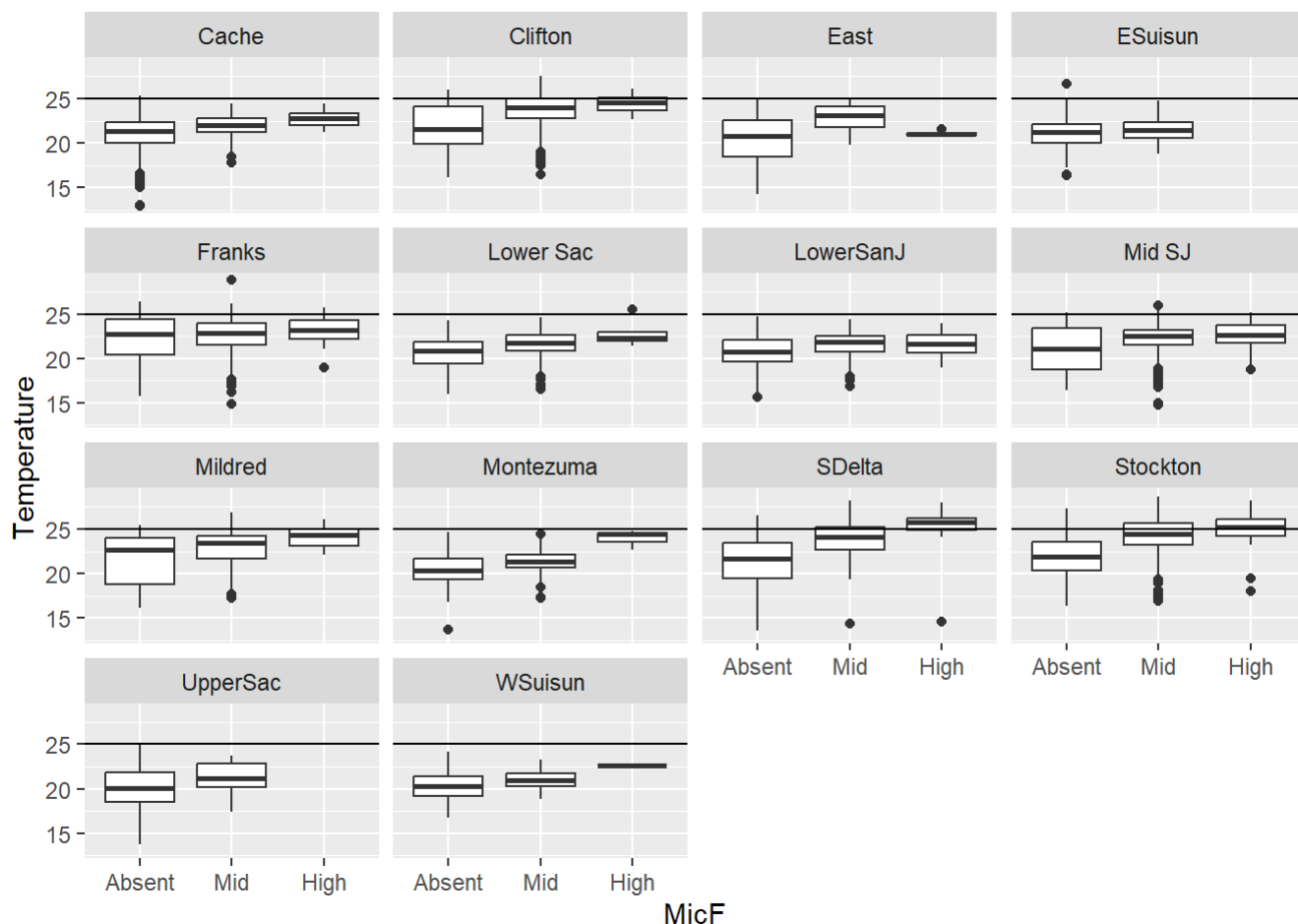


```
ggplot(HABres2, aes(x = Temperature, y = Microcystis))+ geom_point()+  
  facet_wrap(~Region)
```



```
ggplot(HABres2, aes(x = MicF, y = Temperature))+ geom_boxplot()+
  facet_wrap(~Region)+ geom_hline(yintercept = 25)
```





UGH. Still not great

Binomial presence absence by region

```
micbin = group_by(HABres2, Region, Year, Month) %>%
  summarize(Present = length(Microcystis[which(Microcystis>1)]), Absent = length(Microcystis[whi
ch(Microcystis==1)]),
            ResTime = mean(ResTime), Temperature = mean(Temperature, na.rm = T), Salinity = mean
(Salinity, na.rm = T)) %>%
  ungroup() %>%
  mutate(Year2 = Year-2000, Temp = scale(Temperature), Res = scale(ResTime), Sal = sca
le(Salinity))
```

```
## `summarise()` has grouped output by 'Region', 'Year'. You can override using
## the `.groups` argument.
```

```
library(MuMIn)
```

```
micbin2017 = filter(micbin, !is.na(Sal), Year > 2016) %>%
  mutate(logres = log(Res+1))
global17 = glmer(cbind(Present, Absent)~ Temp + logres + Month+ Sal+(1|Year2)+ (1|Region), data
= micbin2017, family = "binomial", na.action = "na.fail")
mods2017 = dredge(global17)
```

```
## Fixed term is "(Intercept)"
```

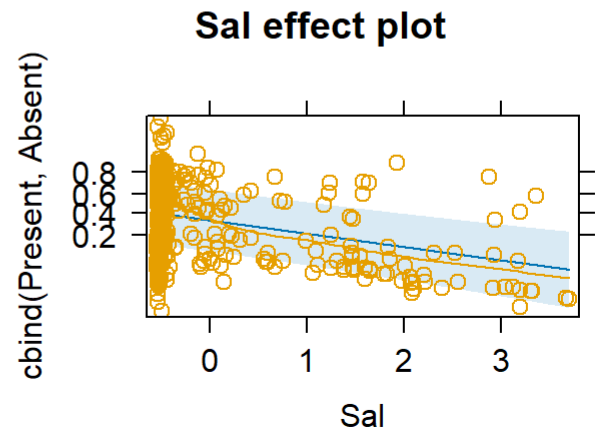
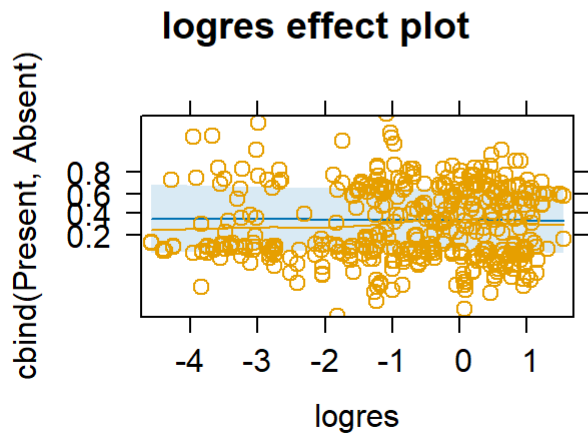
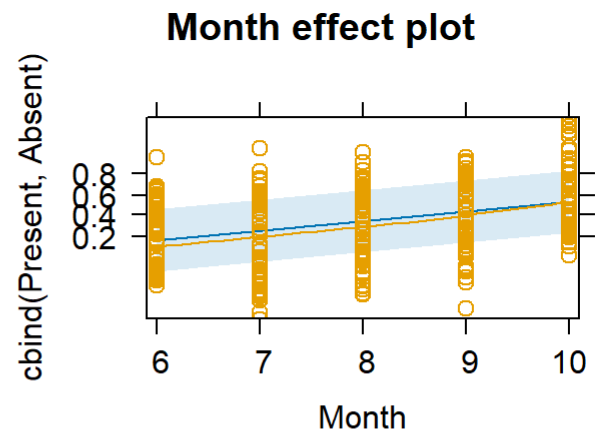
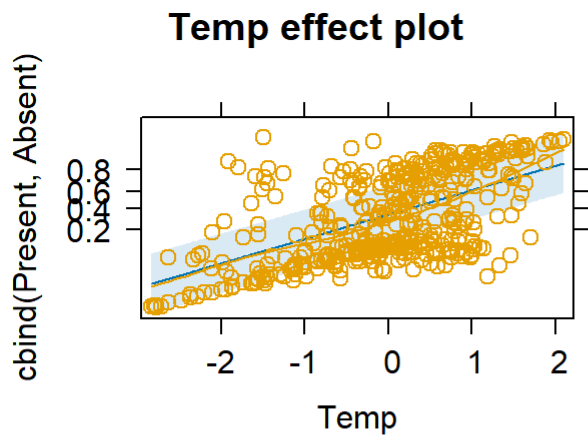
```
write.csv(as.data.frame(mods2017), "MicrocystisModels.csv")
```

```
#Month, salinity, and temperature are the best predictors, residence time close behind
```

```
best17.2 = glmer(cbind(Present, Absent)~ Temp + Month+ logres+Sal+ (1|Year2)+ (1|Region), data
= micbin2017, family = "binomial", na.action = "na.fail")
summary(best17.2)
```

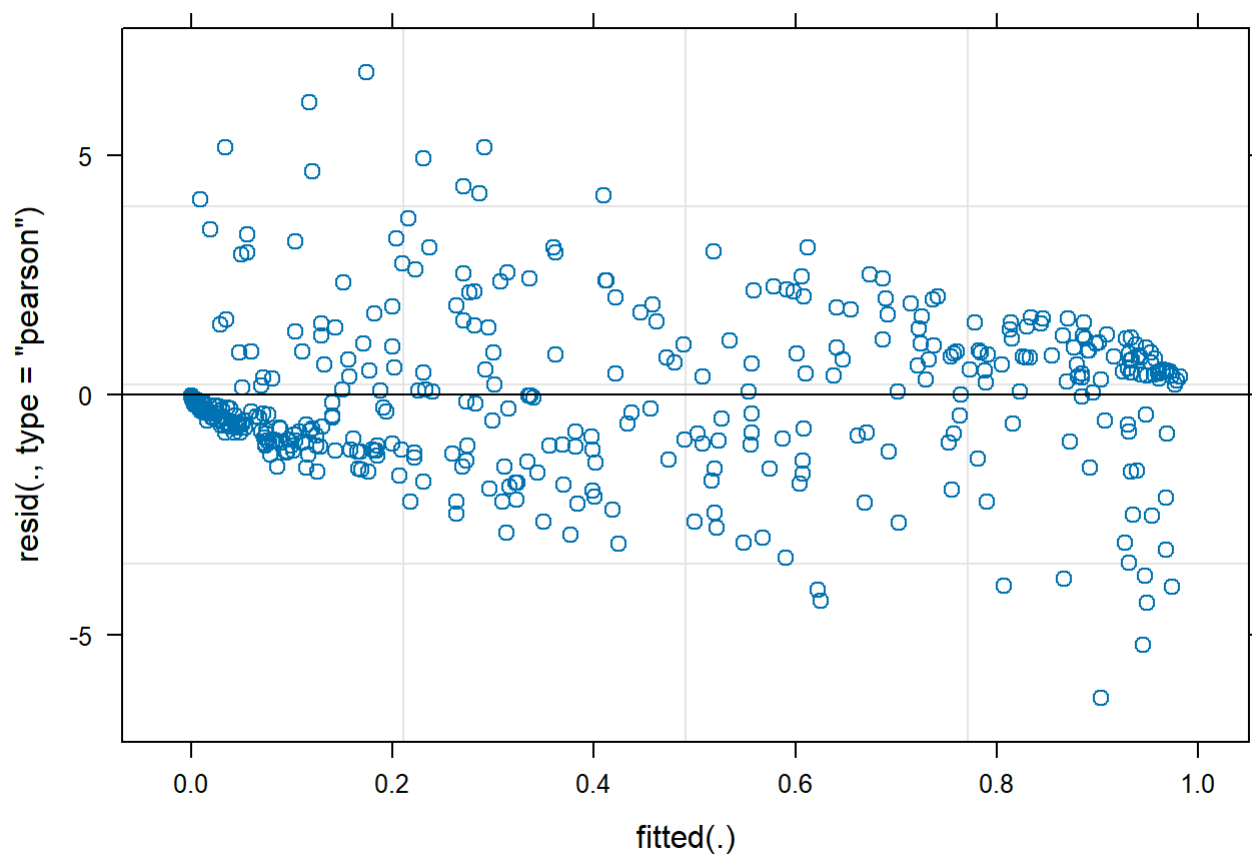
```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: cbind(Present, Absent) ~ Temp + Month + logres + Sal + (1 | Year2) +
## (1 | Region)
## Data: micbin2017
##
##      AIC      BIC   logLik deviance df.resid
## 1747.4   1775.7   -866.7   1733.4     410
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.3253 -0.9828 -0.1827  0.8813  6.7363
##
## Random effects:
## Groups Name      Variance Std.Dev.
## Region (Intercept) 2.402    1.55
## Year2 (Intercept) 1.769    1.33
## Number of obs: 417, groups: Region, 14; Year2, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.08761    0.76999  -5.309 1.10e-07 ***
## Temp         1.12780    0.06899  16.347 < 2e-16 ***
## Month        0.41879    0.04094  10.230 < 2e-16 ***
## logres       -0.01381    0.08620  -0.160  0.873
## Sal          -0.57849    0.13113  -4.412 1.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) Temp   Month  logres
## Temp  -0.262
## Month -0.447  0.595
## logres  0.141  0.007 -0.125
## Sal    0.157 -0.039 -0.270  0.360
```

```
plot(allEffects(best17.2, residuals = TRUE))
```



Let's be sure our diagnostic plots look good before anything else

```
plot(best17.2)
```

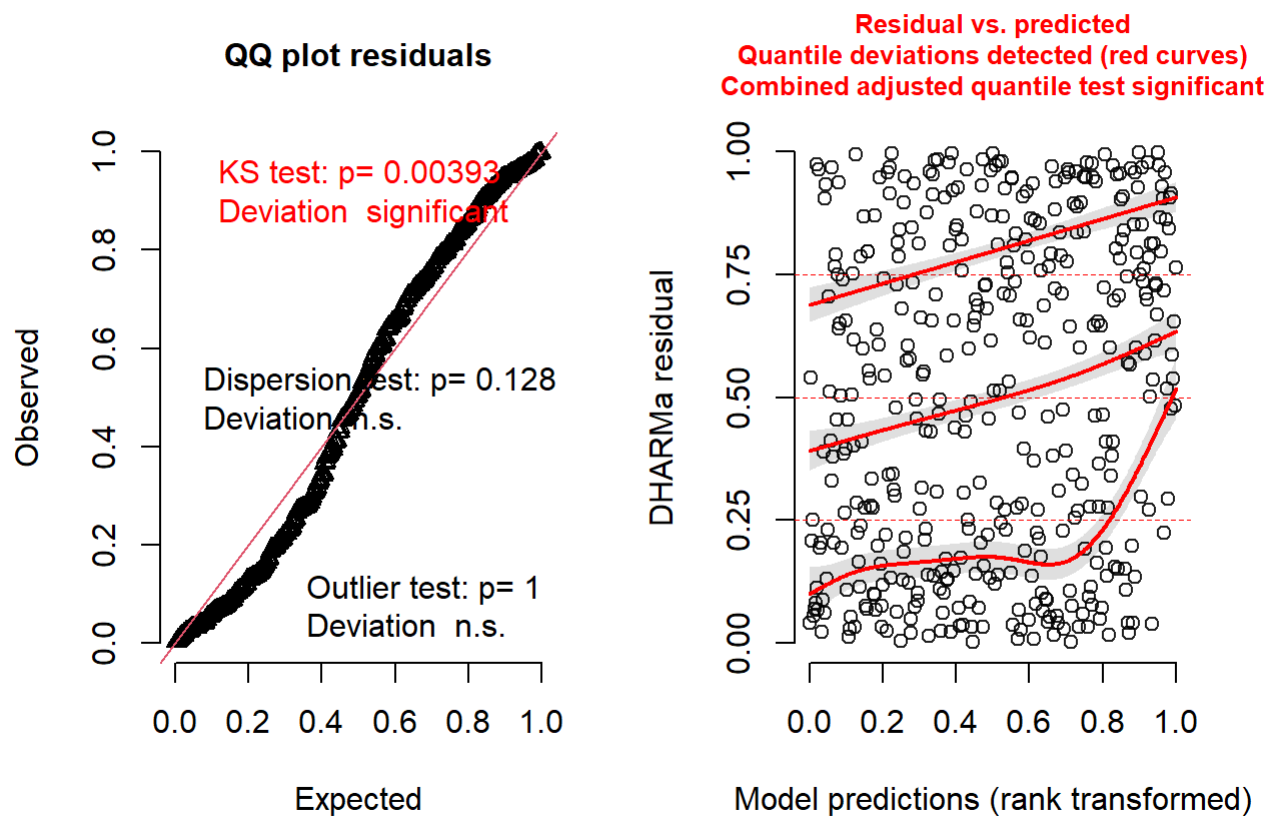


```
library(DHARMA)
```

```
## This is DHARMA 0.4.6. For overview type '?DHARMA'. For recent changes, type news(package = 'DHARMA')
```

```
simres = simulateResiduals(best17.2)  
plot(simres)
```

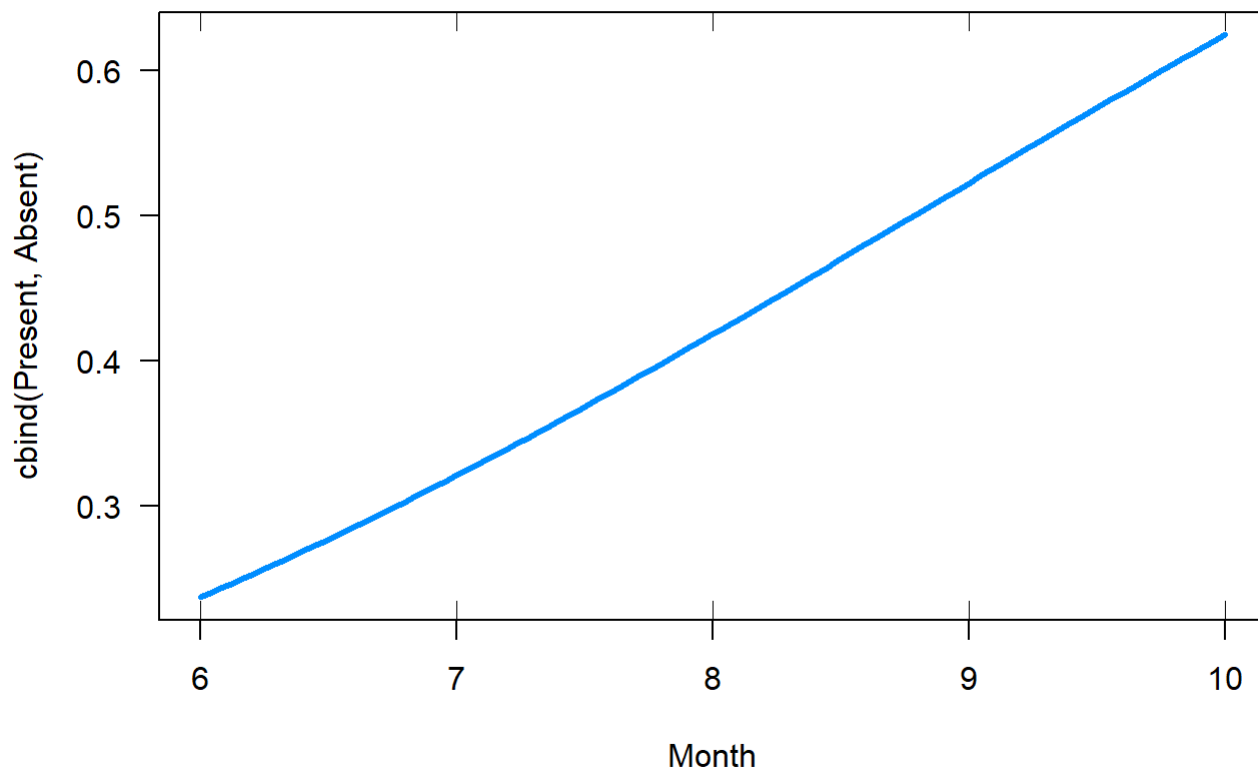
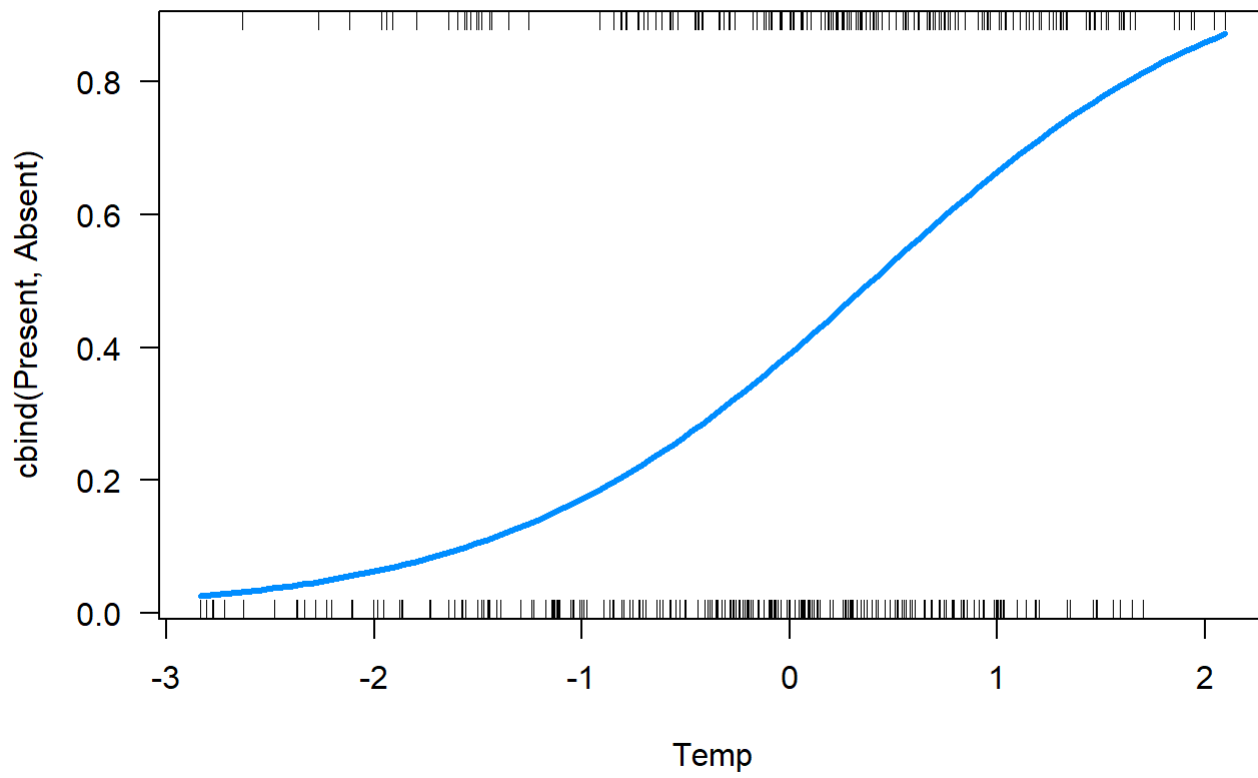
## DHARMa residual

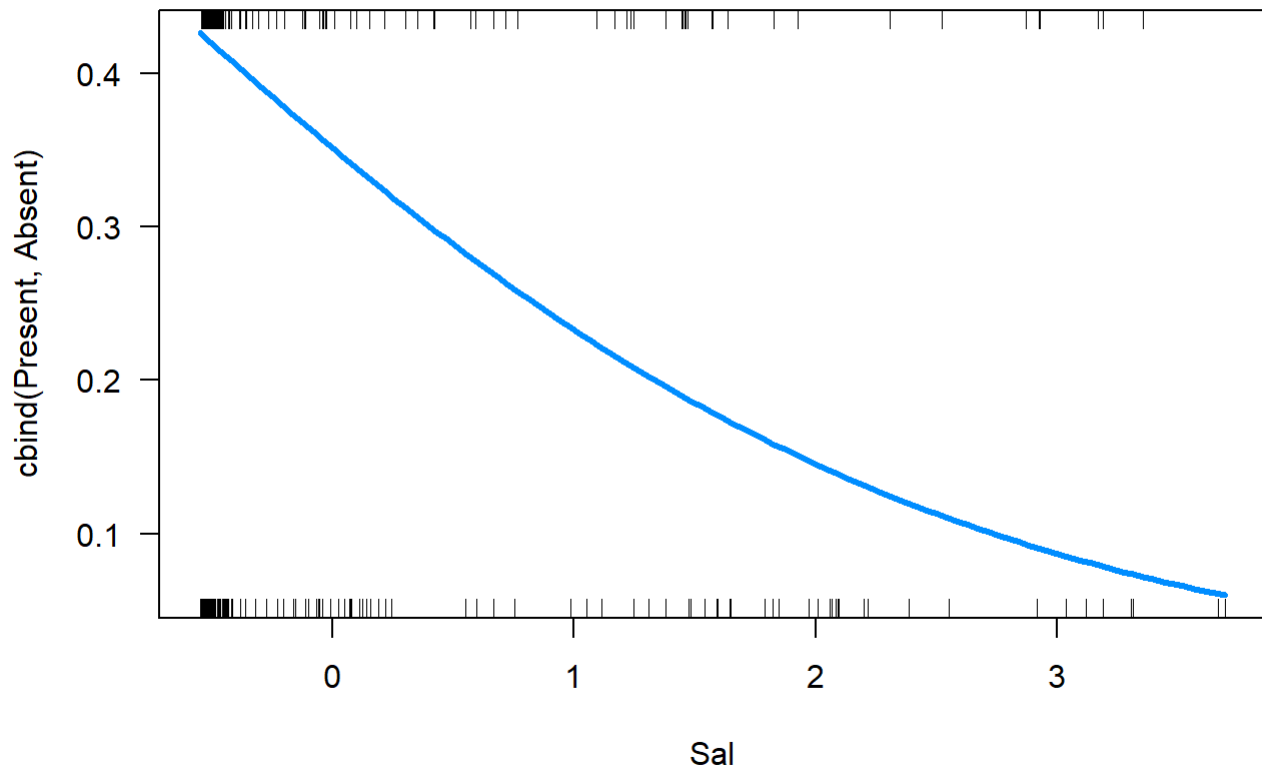
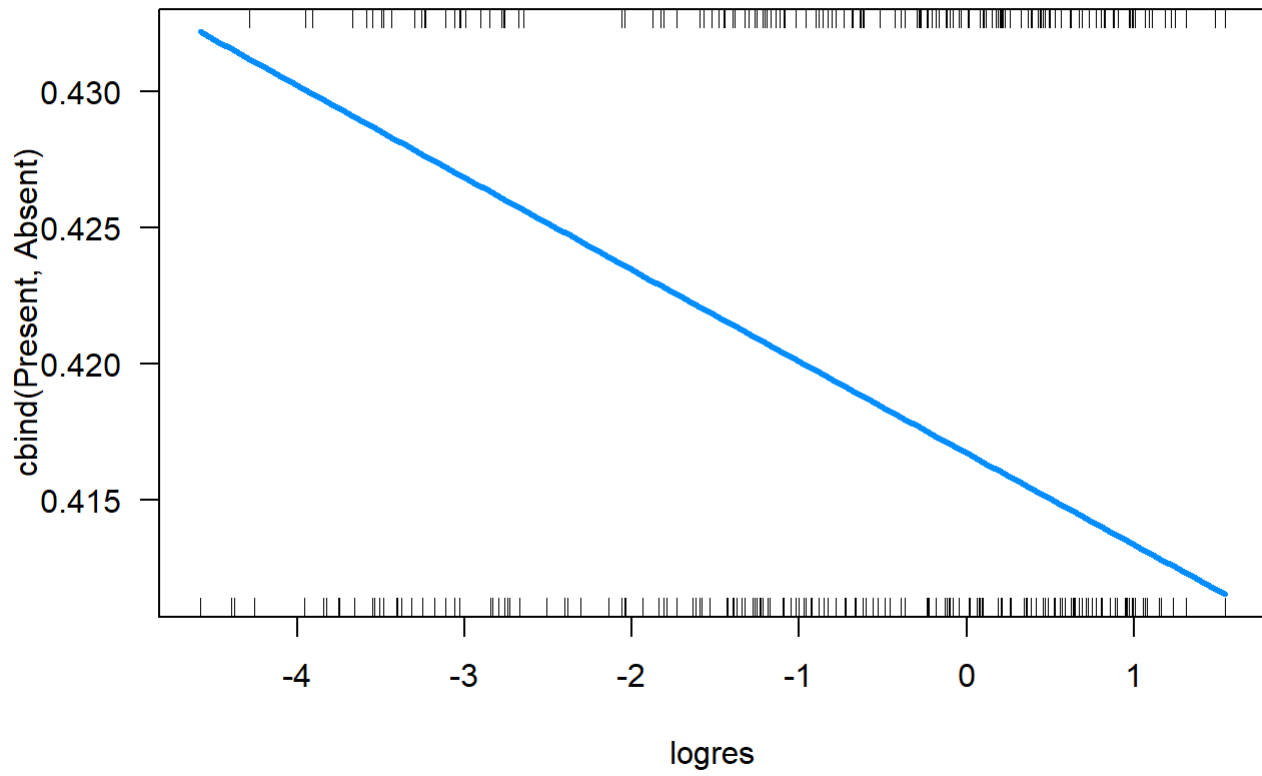


*#not too bad, though some deviation*

```
effs = allEffects(best17.2, partial.residuals = TRUE)
test = ggeffect(best17.2, terms = "Temp")
test2 = ggpredict(best17.2, terms = "Temp [all]")
#OK, need to find a pretty way to plot this
library(visreg)

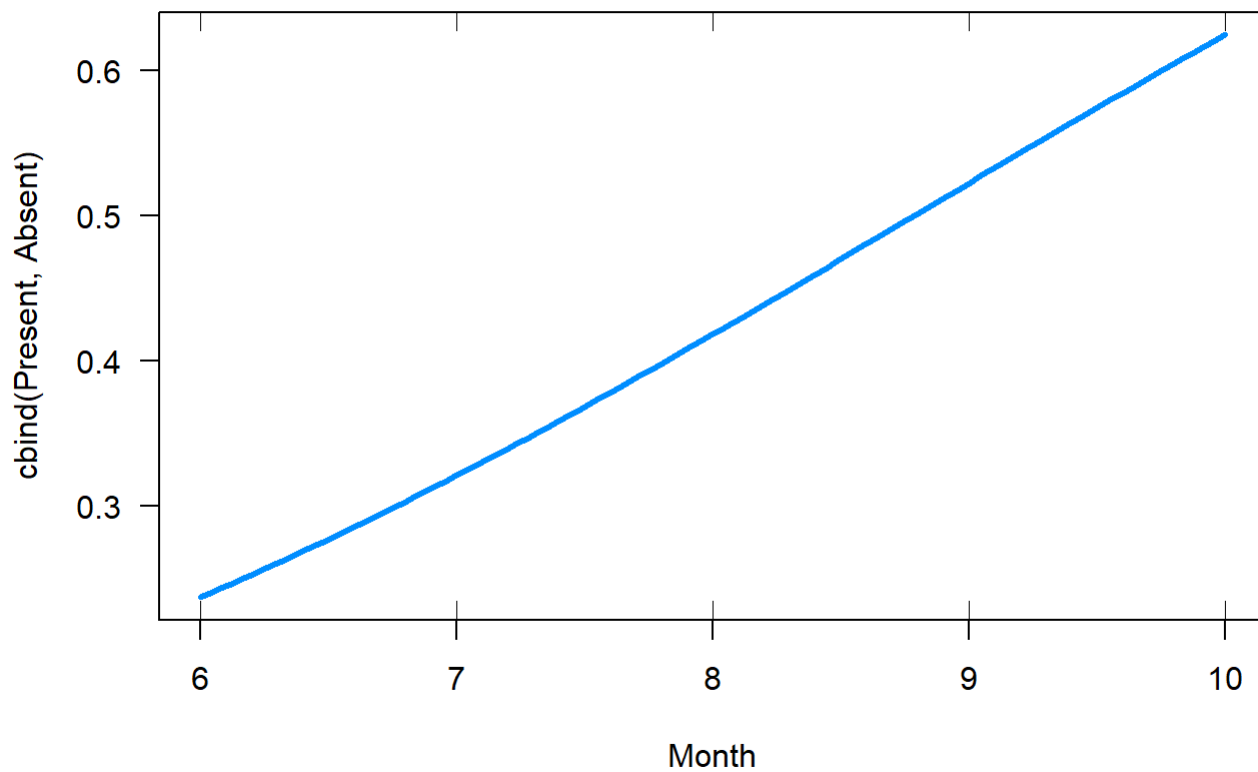
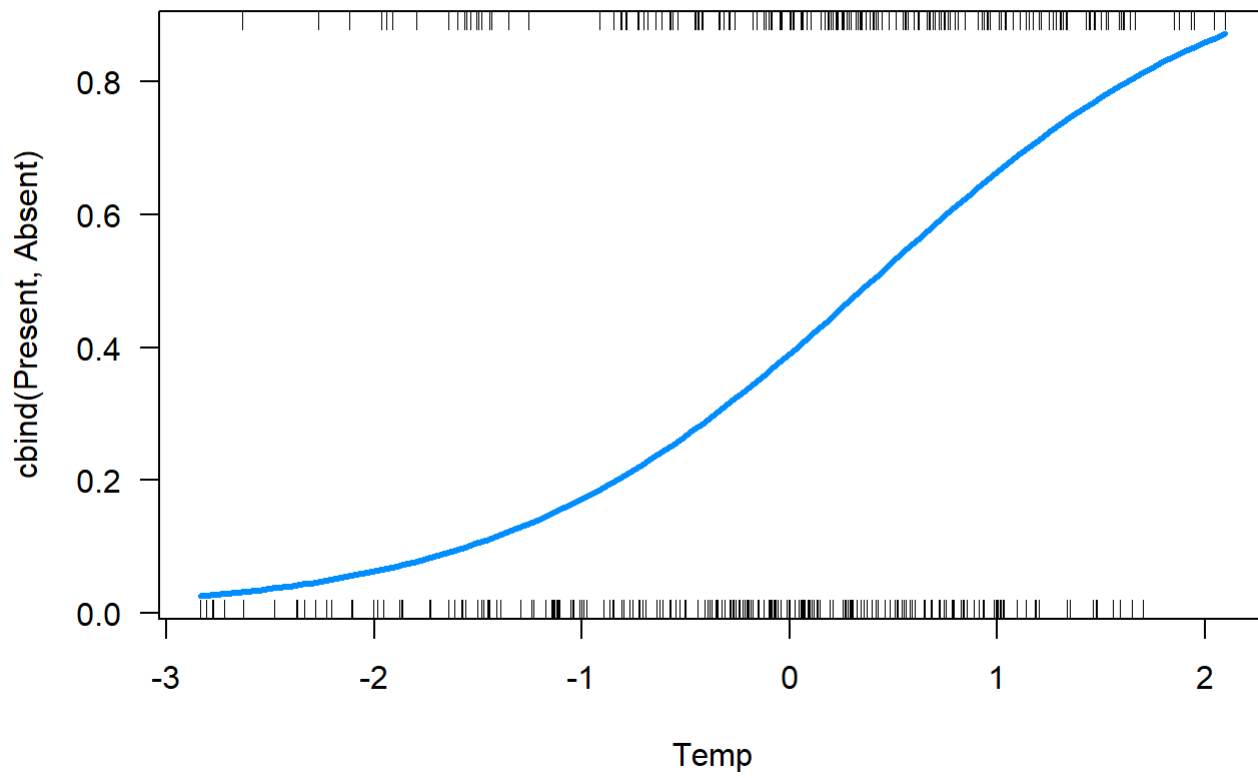
visreg(best17.2, scale = "response")
```

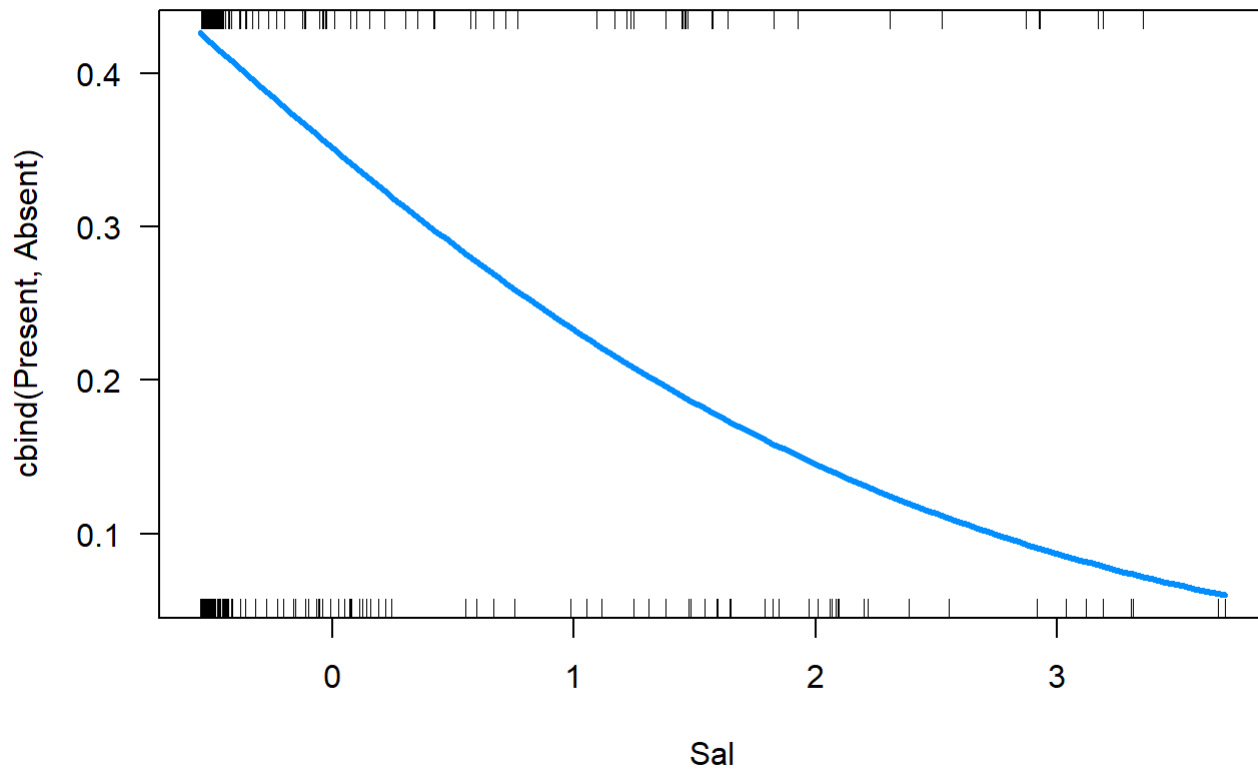
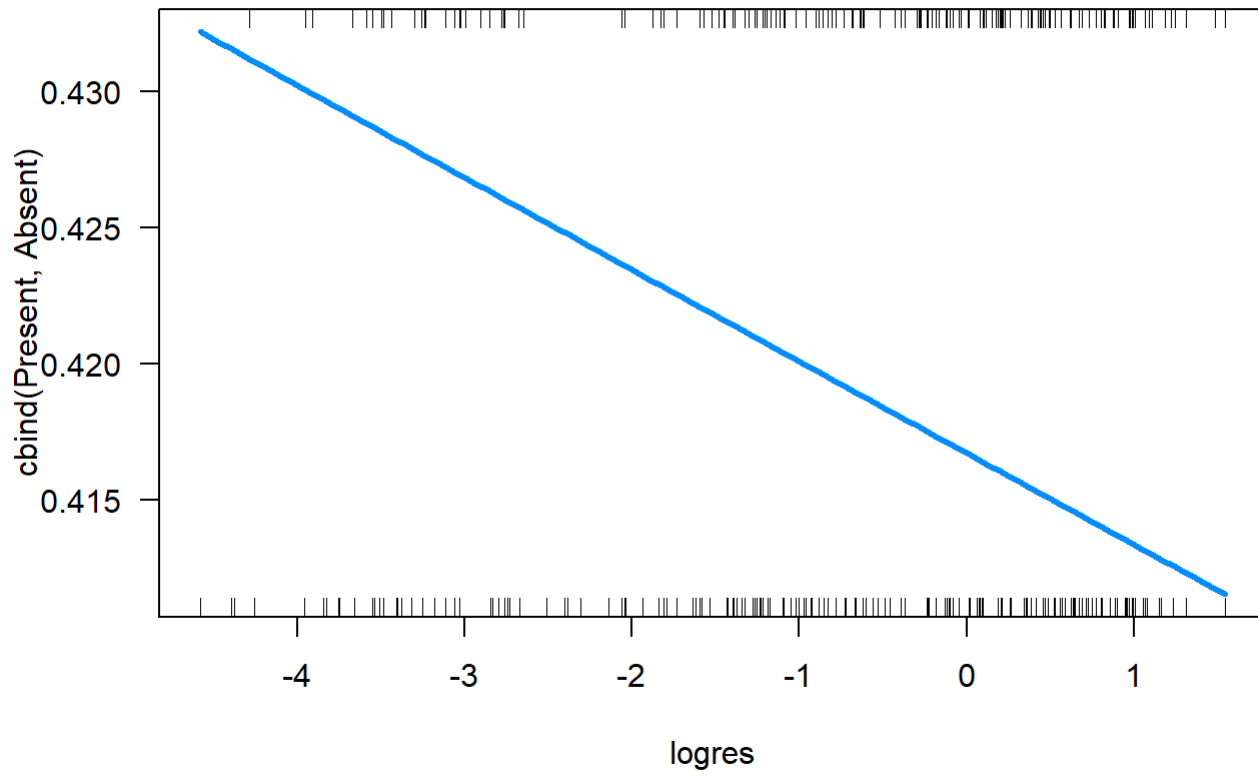




```
foo = visreg(best17.2, scale = "response")
```





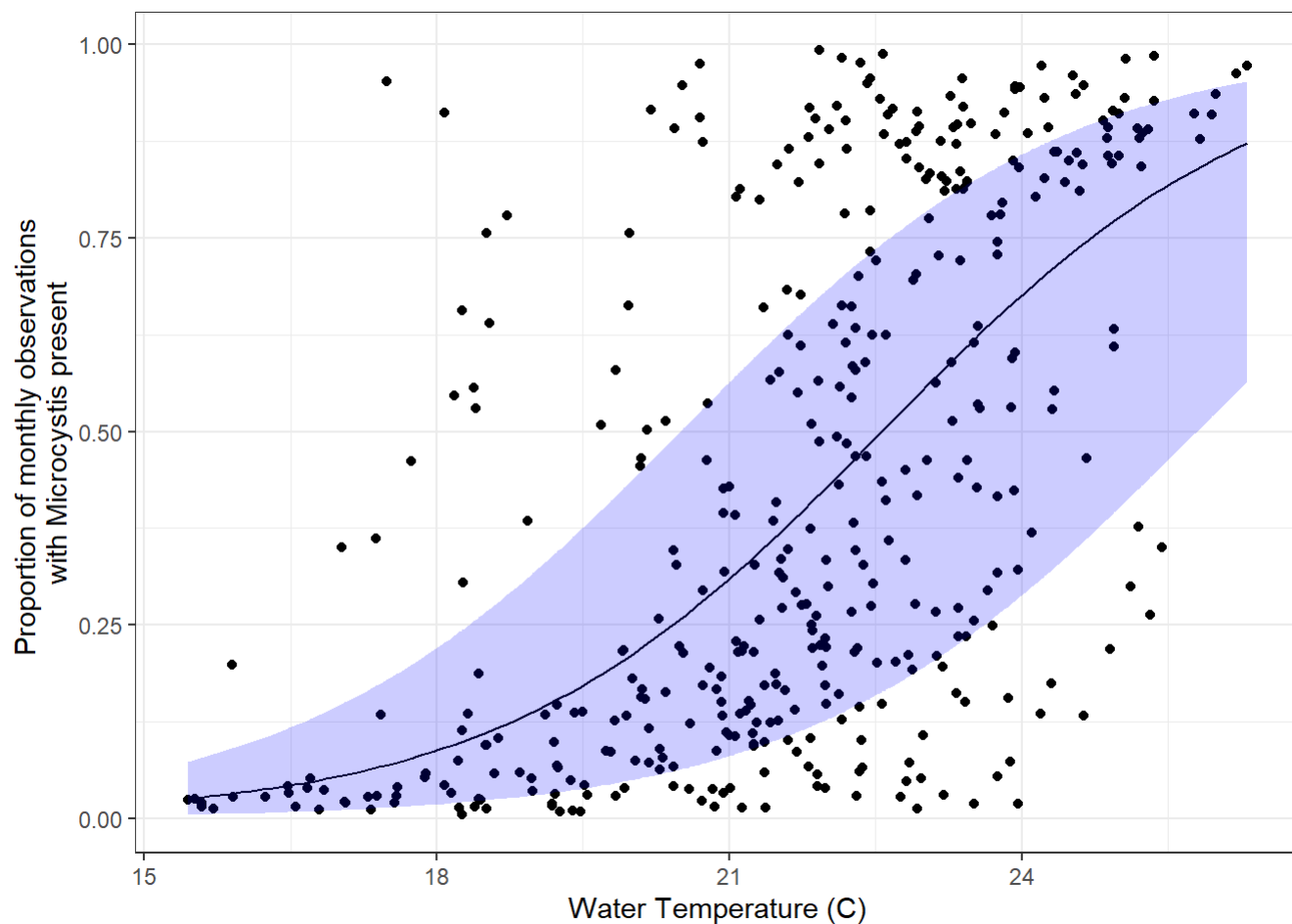


```
Backscale = function(X, dfvar){
  y = X* sd((dfvar-mean(dfvar, na.rm =T)), na.rm =T) + mean(dfvar, na.rm =T)
  return(y)
}
```

```
saltest = lm(Sal ~ Salinity, data = micbin2017)
summary(saltest)
```

```
##
## Call:
## lm(formula = Sal ~ Salinity, data = micbin2017)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.037e-15 -3.610e-17  7.100e-18  2.280e-17  2.204e-15
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -5.592e-01  1.328e-17 -4.209e+16  <2e-16 ***
## Salinity     2.922e-01  3.638e-18  8.030e+16  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.391e-16 on 415 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 6.447e+33 on 1 and 415 DF, p-value: < 2.2e-16
```

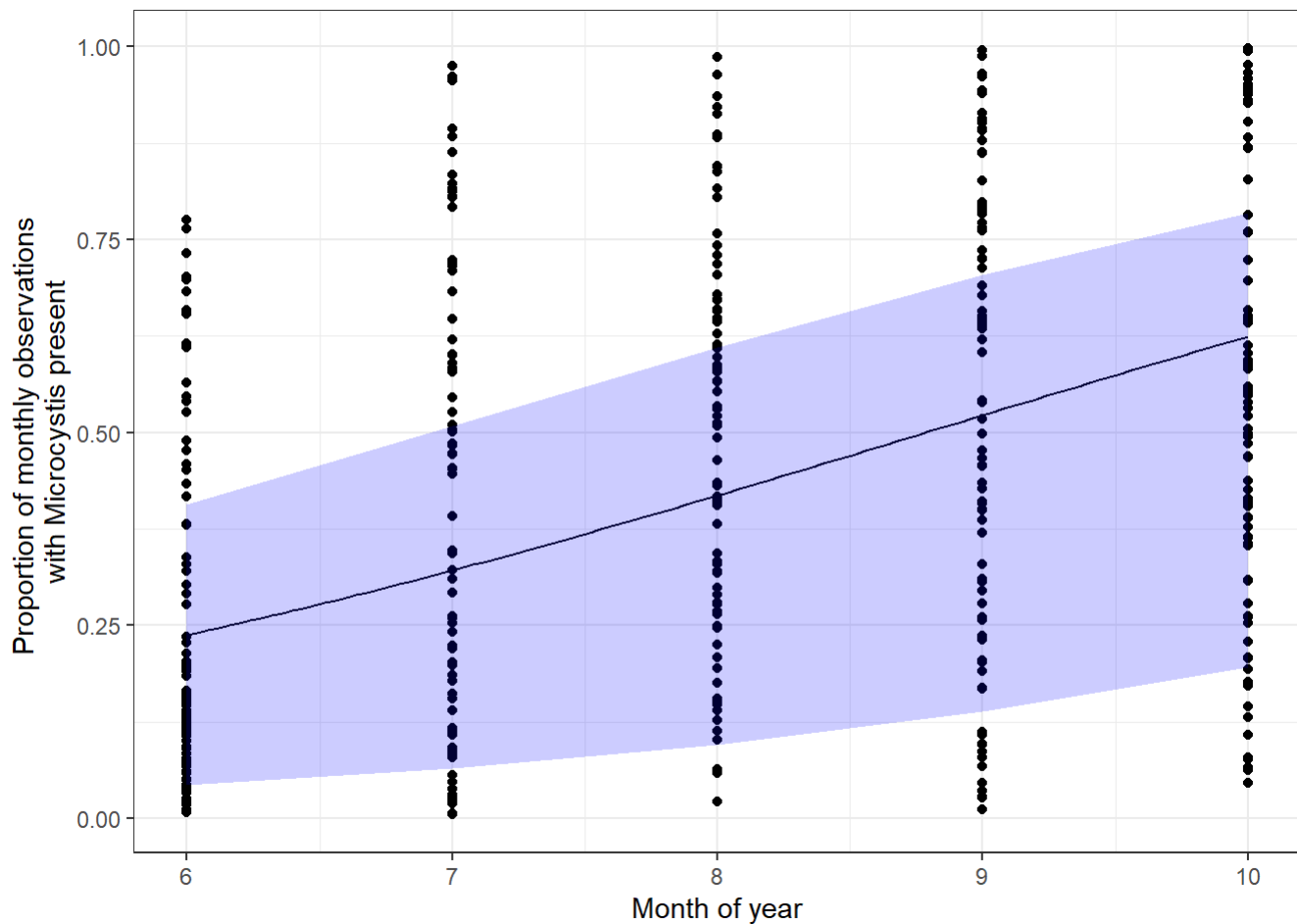
```
#OK, now all the partial residual plots!!
#first temperature
ggplot() + geom_point(data =foo[[1]]$res, aes(x = Backscale(Temp, micbin2017$Temperature), y = v
isregRes))+
  geom_line(data = foo[[1]]$fit, aes(x = Backscale(Temp, micbin2017$Temperature), y = visregFi
t))+
  geom_ribbon(data = test2, aes(x = Backscale(x, micbin2017$Temperature), ymin = conf.low, ymax
= conf.high),
            alpha = 0.2, fill = "blue")+
  theme_bw()+
  ylab("Proportion of monthly observations \nwith Microcystis present")+
  xlab("Water Temperature (C)")
```



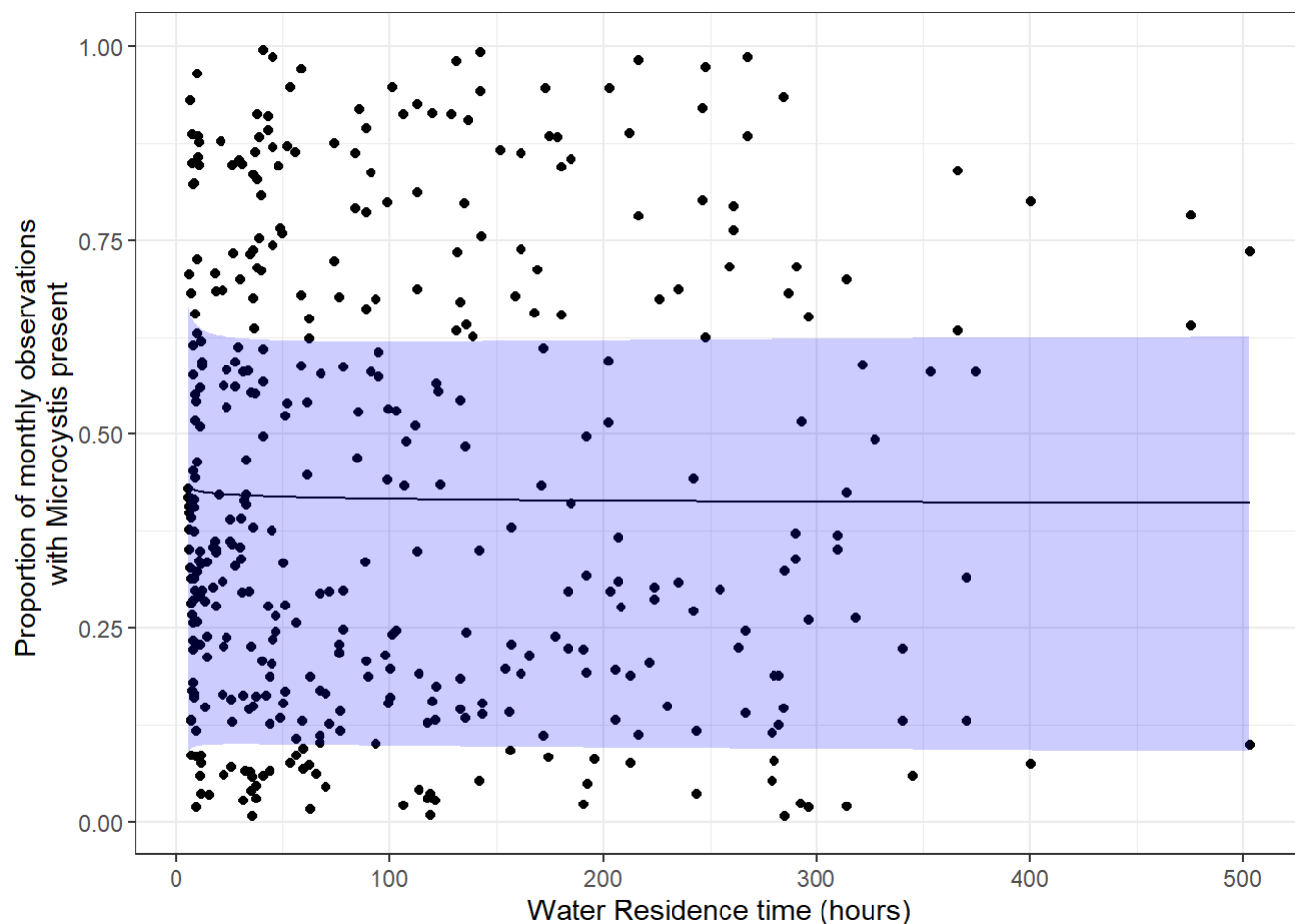
```
ggsave("plots/Templot.tiff", device = "tiff", width = 5, height = 5)
```

```
#Now month
```

```
montheff = ggpredict(best17.2, terms = "Month [all]")
ggplot() + geom_point(data = foo[[2]]$res, aes(x = Month, y = visregRes))+
  geom_line(data = foo[[2]]$fit, aes(x = Month, y = visregFit))+
  geom_ribbon(data = montheff, aes(x = x, ymin = conf.low, ymax = conf.high),
    alpha = 0.2, fill = "blue")+
  theme_bw()+
  ylab("Proportion of monthly observations \nwith Microcystis present")+
  xlab("Month of year")
```



```
#residnece time
reseff = ggpredict(best17.2, terms = "logres [all]")
ggplot() + geom_point(data = foo[[3]]$res, aes(x = Backscale(exp(logres)-1, micbin2017$ResTime), y = visregRes)) +
  geom_line(data = foo[[3]]$fit, aes(x = Backscale(exp(logres)-1, micbin2017$ResTime), y = visregFit)) +
  geom_ribbon(data = reseff, aes(x = Backscale(exp(x)-1, micbin2017$ResTime), ymin = conf.low, ymax = conf.high),
    alpha = 0.2, fill = "blue") +
  theme_bw() +
  ylab("Proportion of monthly observations \nwith Microcystis present") +
  xlab("Water Residence time (hours)")
```



```
ggsave("plots/Resplot.tiff", device = "tiff", width = 5, height = 5)
```

```
#Salinity
```

```
saleff = ggpredict(best17.2, terms = "Sal [all]")
```

```
ggplot() + geom_point(data = foo[[4]]$res, aes(x = Backscale(Sal, micbin2017$Salinity), y = visreg  
Res)) +
```

```
  geom_line(data = foo[[4]]$fit, aes(x = Backscale(Sal, micbin2017$Salinity), y = visregFit)) +
```

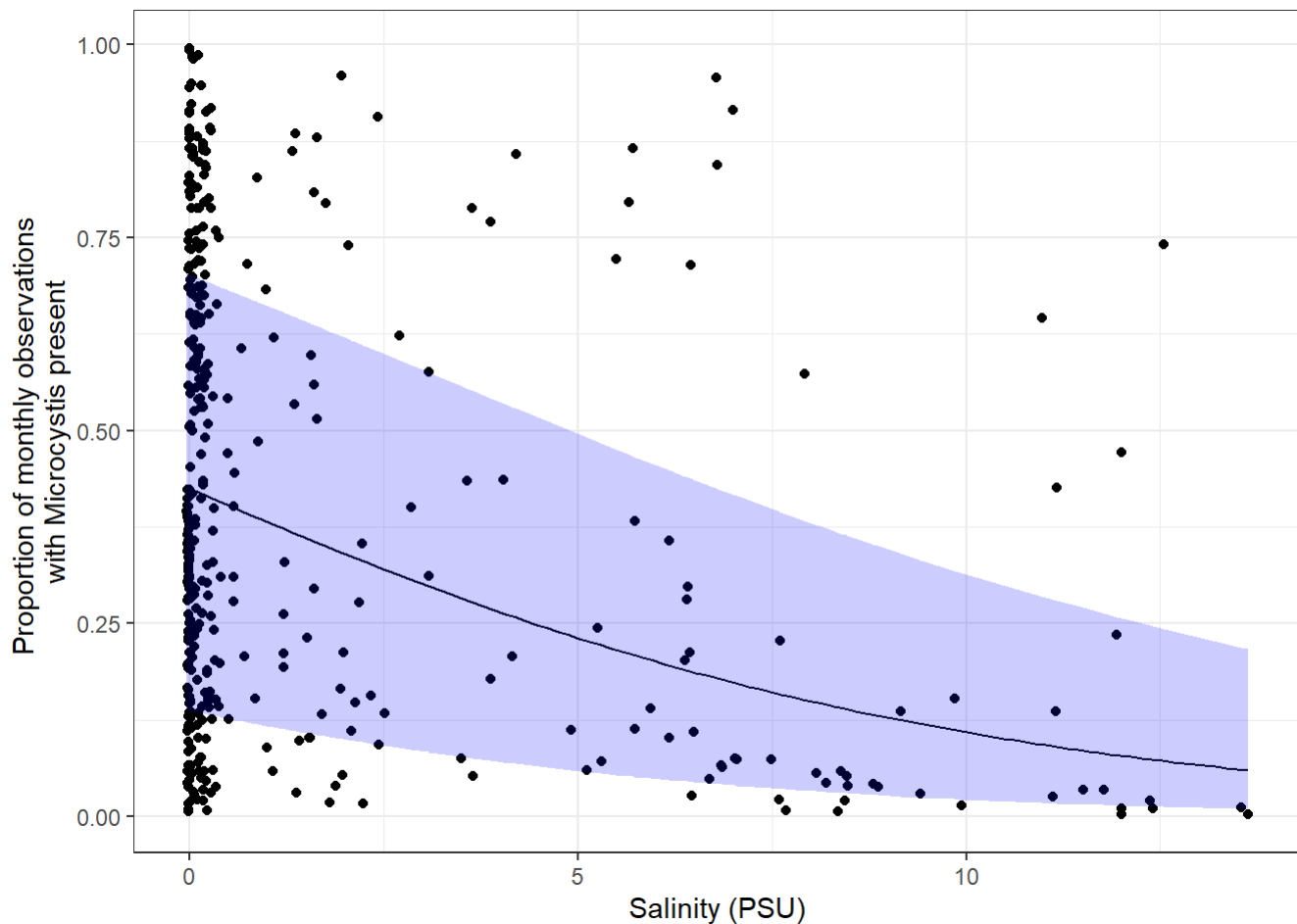
```
  geom_ribbon(data = saleff, aes(x = Backscale(x, micbin2017$Salinity), ymin = conf.low, ymax =  
conf.high),
```

```
    alpha = 0.2, fill = "blue") +
```

```
  theme_bw() +
```

```
  ylab("Proportion of monthly observations \nwith Microcystis present") +
```

```
  xlab("Salinity (PSU)")
```



```
ggsave("plots/Salplot.tiff", device = "tiff", width = 5, height = 5)
```

```
#Pretty version of model results
library(broom.mixed)
summod = tidy(best17.2)

write.csv(summod, "Micmod.csv")
```

Mine wants a more balanced design

```
micbin2 = filter(HABres2, Year > 2016, Source != "Prop1", Source != "StocktonWF") %>%
  group_by(Region, Year, Month) %>%
  summarize(Present = length(Microcystis[which(Microcystis>1)]), Absent = length(Microcystis[whi
ch(Microcystis==1)]),
            ResTime = mean(ResTime), Temperature = mean(Temperature, na.rm = T), Salinity = mean
(Salinity, na.rm = T)) %>%
  ungroup() %>%
  mutate(Year2 = Year-2000, Temp = scale(Temperature), Res = scale(ResTime), Sal = sca
le(Salinity),
         logres = log(Res +1))
```

```
## `summarise()` has grouped output by 'Region', 'Year'. You can override using
## the `.groups` argument.
```

```
global2 = glmer(cbind(Present, Absent)~ Temp + logres + Month+ Sal+(1|Year2)+ (1|Region), data =  
micbin2, family = "binomial", na.action = "na.fail")  
mods2 = dredge(global2)
```

```
## Fixed term is "(Intercept)"
```

```
write.csv(as.data.frame(mods2), "MicrocystisModels2.csv")
```

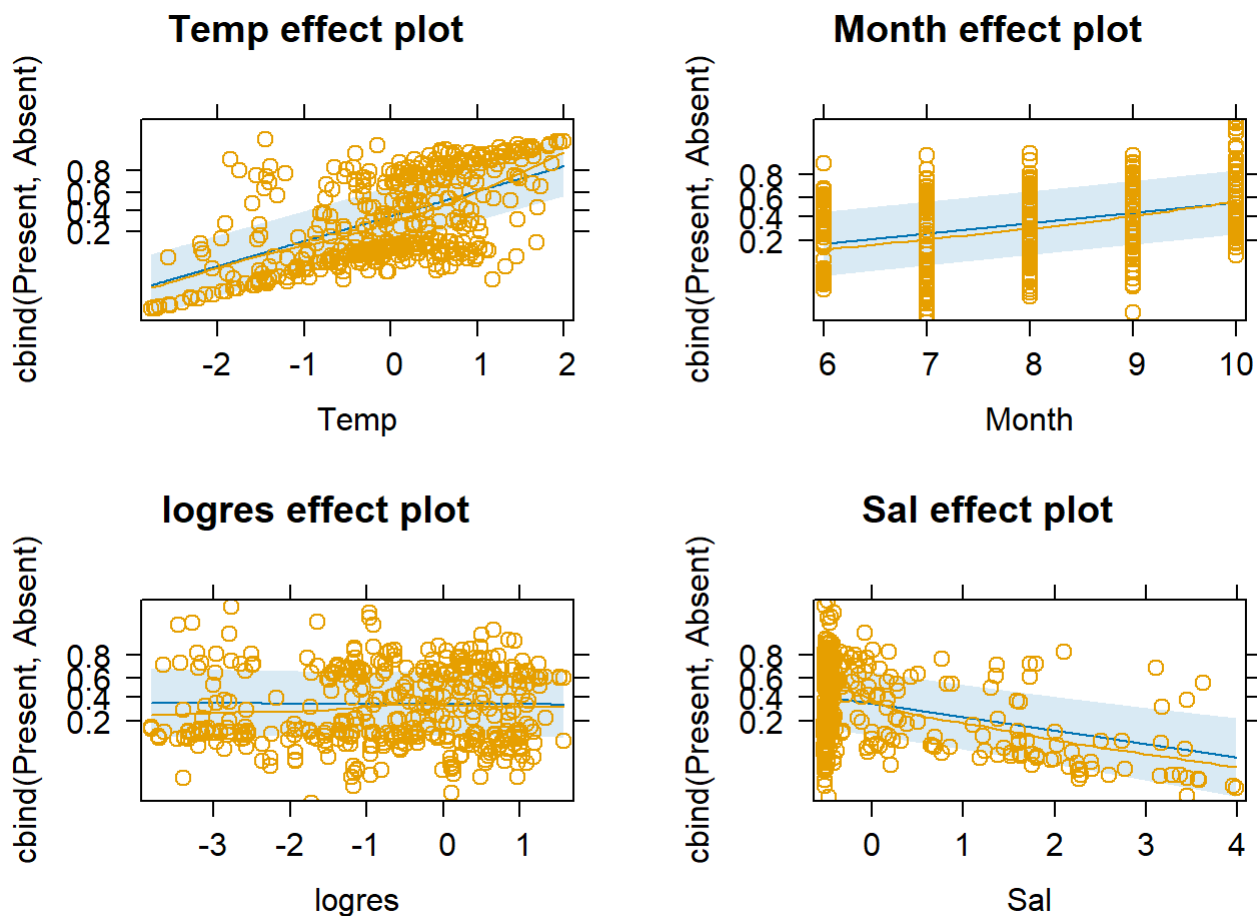
*#Month, salinity, and temperature are the best predictors, residence time close behind*

```
best2 = glmer(cbind(Present, Absent)~ Temp + Month+ logres+Sal+ (1|Year2)+ (1|Region), data = m  
icbin2, family = "binomial", na.action = "na.fail")  
summary(best2)
```



```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: cbind(Present, Absent) ~ Temp + Month + logres + Sal + (1 | Year2) +
## (1 | Region)
## Data: micbin2
##
##      AIC      BIC   logLik deviance df.resid
## 1689.1   1717.4   -837.6   1675.1     410
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.3287 -0.9300 -0.1419  0.8788  6.5175
##
## Random effects:
## Groups Name      Variance Std.Dev.
## Region (Intercept) 2.264    1.505
## Year2 (Intercept) 1.818    1.348
## Number of obs: 417, groups: Region, 14; Year2, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.18967    0.77332  -5.418 6.04e-08 ***
## Temp         1.13407    0.07257  15.627 < 2e-16 ***
## Month        0.43760    0.04258  10.276 < 2e-16 ***
## logres       -0.01120    0.09014  -0.124  0.901
## Sal         -0.56920    0.12480  -4.561 5.10e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) Temp   Month logres
## Temp  -0.272
## Month -0.459  0.598
## logres  0.128  0.014 -0.114
## Sal    0.151 -0.040 -0.275  0.362
```

```
plot(allEffects(best2, residuals = TRUE))
```



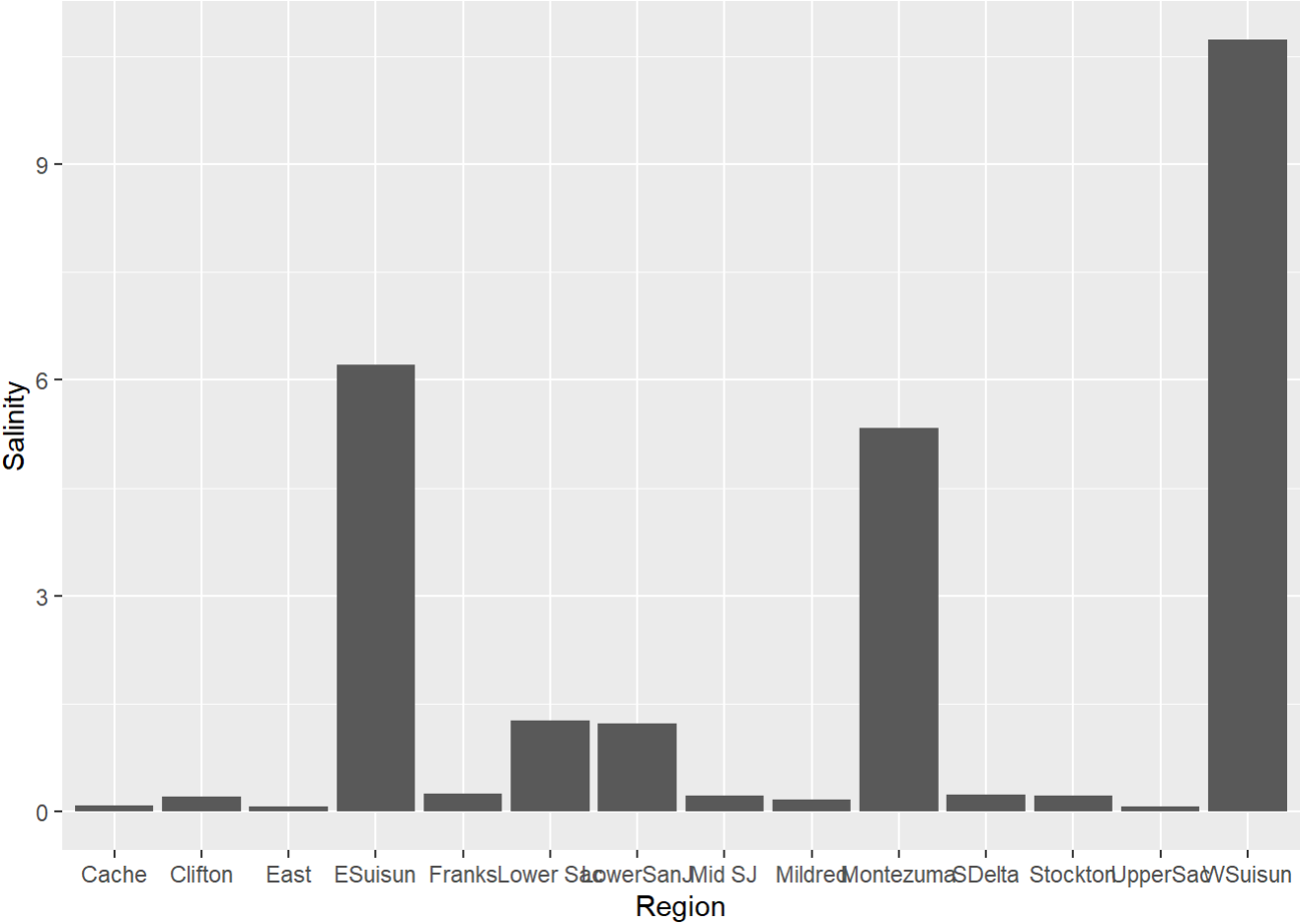
*#pretty much the same result*

Look at data by region, compare physical variables.

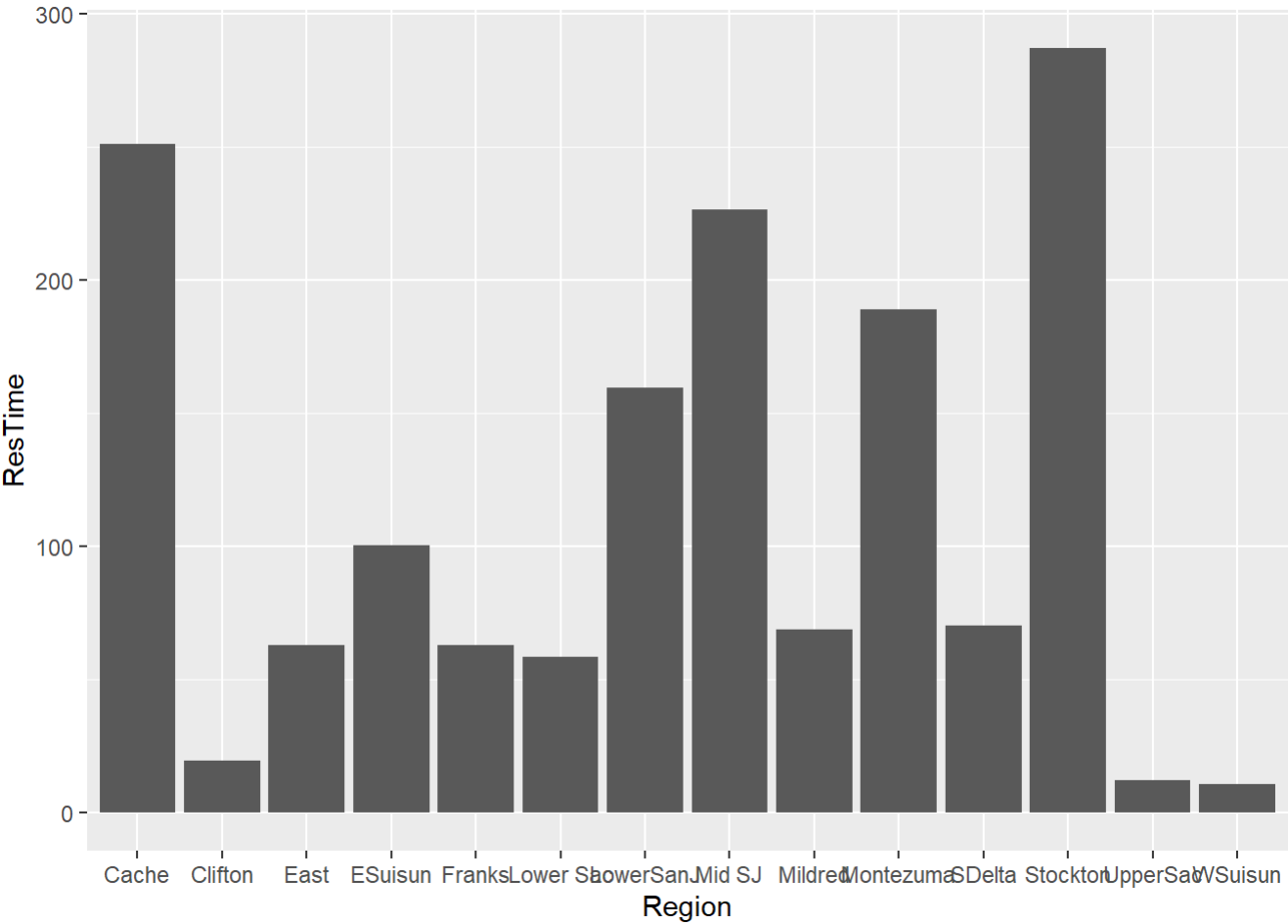
```
physreg = group_by(HABres2, Region, Year) %>%
  summarize(Present = length(Microcystis[which(Microcystis>1)]), Absent = length(Microcystis[whi
ch(Microcystis==1)]), Percent = Present/(Present+Absent),
            ResTime = mean(ResTime), Temperature = mean(Temperature, na.rm =T), Salinity = mean
(Salinity, na.rm = T),
            Secchi = mean(Secchi, na.rm =T), Depth = mean(Depth, na.rm =T)) %>%
  group_by(Region) %>%
  summarise(Present = mean(Present), Absent = mean(Absent), Percent =Present/(Present+Absent), R
esTime = mean(ResTime), Temperature = mean(Temperature, na.rm =T), Salinity = mean(Salinity, na.
rm = T), Secchi = mean(Secchi, na.rm =T), Depth = mean(Depth, na.rm =T))
```

## `summarise()` has grouped output by 'Region'. You can override using the  
## `.groups` argument.

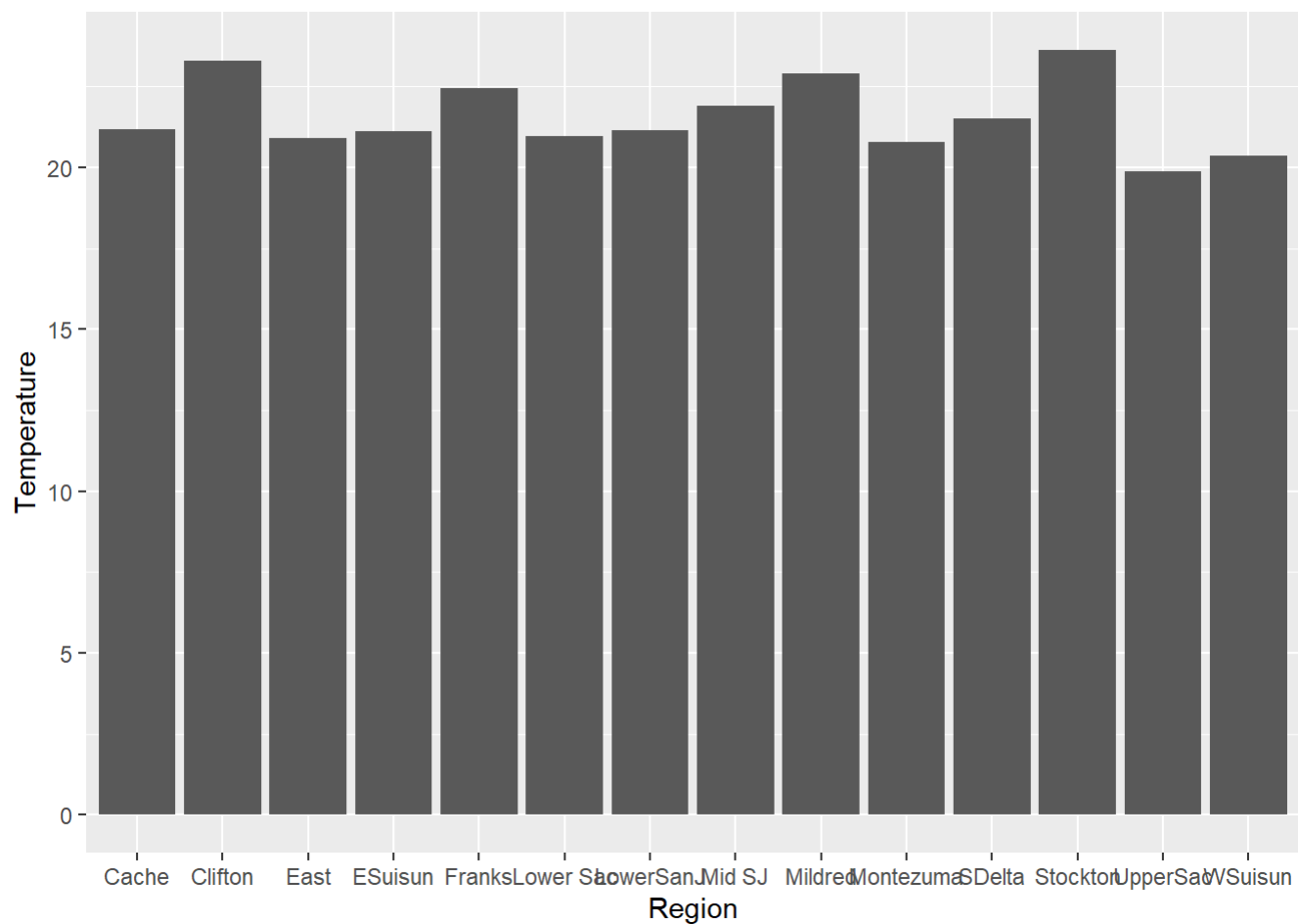
```
ggplot(physreg, aes(x = Region, y = Salinity))+ geom_col()
```



```
ggplot(physreg, aes(x = Region, y = ResTime))+ geom_col()
```

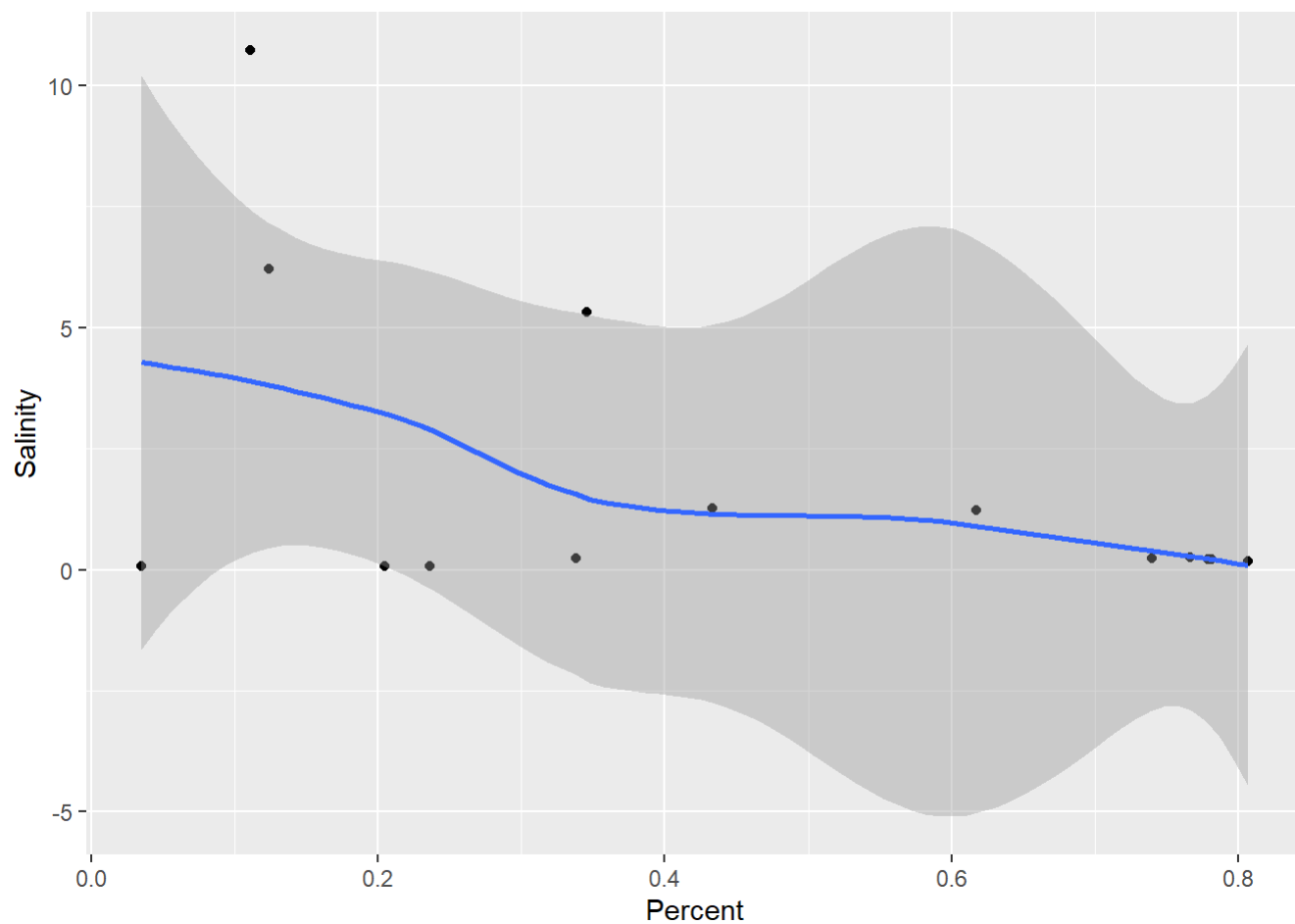


```
ggplot(physreg, aes(x = Region, y = Temperature))+ geom_col()
```



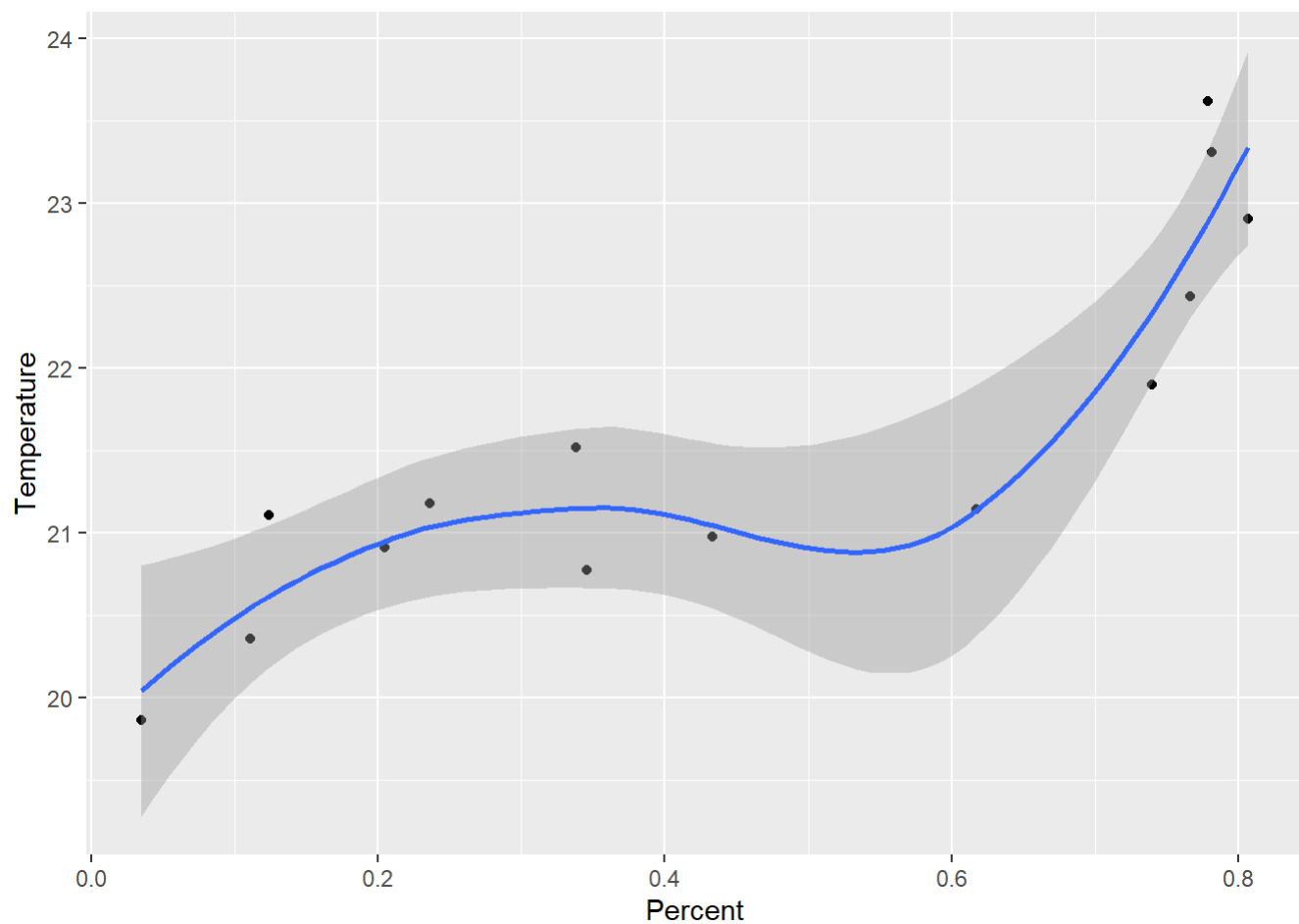
```
ggplot(physreg, aes(x = Percent, y = Salinity))+ geom_point()+ geom_smooth()
```

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



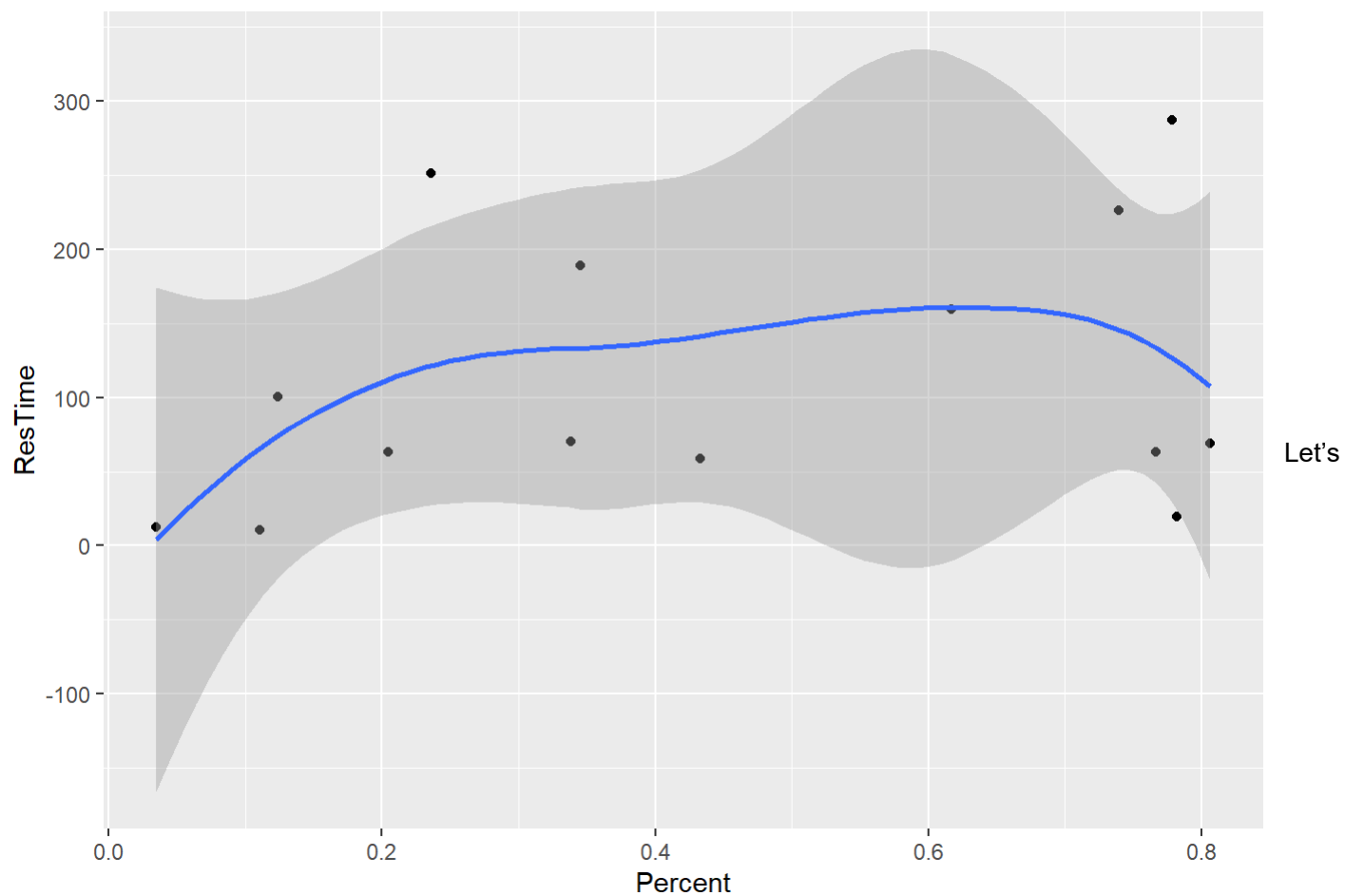
```
ggplot(physreg, aes(x = Percent, y = Temperature))+ geom_point()+ geom_smooth()
```

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



```
ggplot(physreg, aes(x = Percent, y = ResTime))+ geom_point()+ geom_smooth()
```

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



try an NMDS plot

```
physreg = mutate(physreg, Depth = case_when(is.nan(Depth) ~ 8,
                                             TRUE ~ Depth))
regmat = as.matrix(select(physreg, Temperature, Salinity, ResTime, Depth, Secchi))

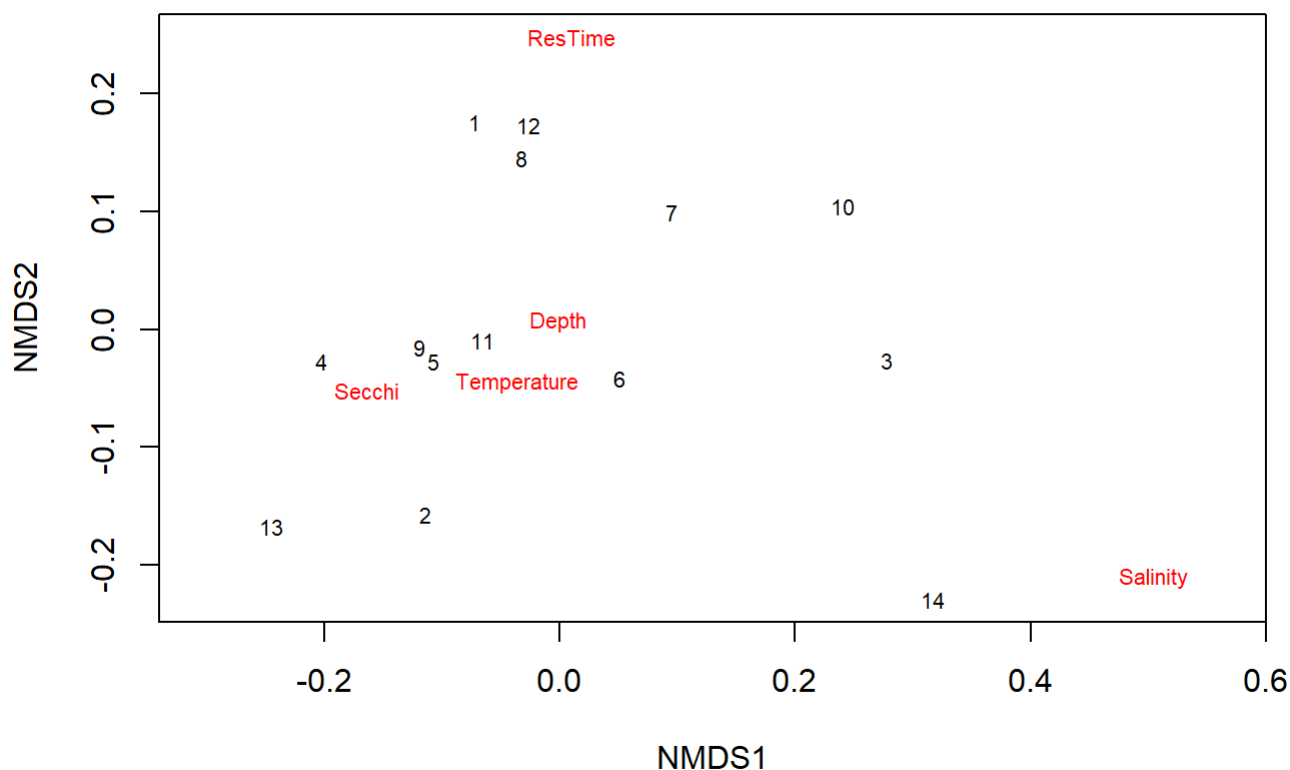
mds1 = metaMDS(regmat, autotransform = T)
```



```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.03715665
## Run 1 stress 0.03715666
## ... Procrustes: rmse 2.646165e-05 max resid 6.39725e-05
## ... Similar to previous best
## Run 2 stress 0.0598041
## Run 3 stress 0.2804498
## Run 4 stress 0.0598041
## Run 5 stress 0.0598041
## Run 6 stress 0.03715665
## ... New best solution
## ... Procrustes: rmse 4.017872e-06 max resid 8.567225e-06
## ... Similar to previous best
## Run 7 stress 0.03715665
## ... New best solution
## ... Procrustes: rmse 9.1917e-06 max resid 2.234403e-05
## ... Similar to previous best
## Run 8 stress 0.03715666
## ... Procrustes: rmse 1.62986e-05 max resid 3.918239e-05
## ... Similar to previous best
## Run 9 stress 0.03715665
## ... New best solution
## ... Procrustes: rmse 4.761761e-06 max resid 1.133015e-05
## ... Similar to previous best
## Run 10 stress 0.03715665
## ... New best solution
## ... Procrustes: rmse 9.299208e-07 max resid 2.06624e-06
## ... Similar to previous best
## Run 11 stress 0.0598041
## Run 12 stress 0.03715665
## ... Procrustes: rmse 6.208699e-06 max resid 1.455219e-05
## ... Similar to previous best
## Run 13 stress 0.03715665
## ... Procrustes: rmse 4.643656e-06 max resid 1.048287e-05
## ... Similar to previous best
## Run 14 stress 0.0598041
## Run 15 stress 0.03715665
## ... Procrustes: rmse 1.630244e-05 max resid 3.881217e-05
## ... Similar to previous best
## Run 16 stress 0.03715665
## ... Procrustes: rmse 6.442109e-06 max resid 1.586081e-05
## ... Similar to previous best
## Run 17 stress 0.2607879
## Run 18 stress 0.0598041
## Run 19 stress 0.03715665
## ... Procrustes: rmse 3.806056e-06 max resid 8.80924e-06
## ... Similar to previous best
## Run 20 stress 0.03715665
## ... Procrustes: rmse 1.158419e-05 max resid 2.515813e-05
```

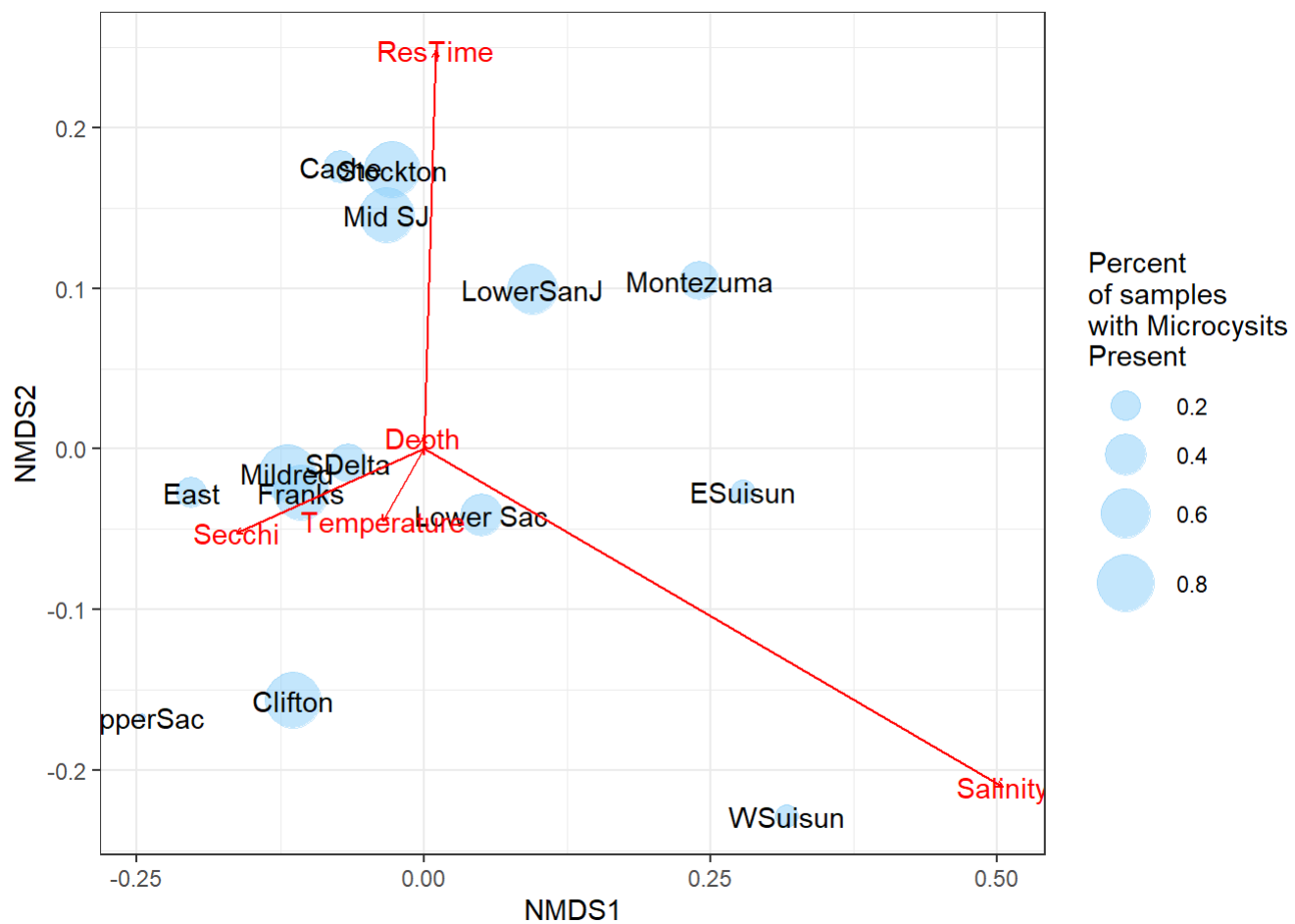
```
## ... Similar to previous best
## *** Best solution repeated 7 times
```

```
plot(mds1, type = "t")
```



```
data.scores <- as.data.frame(scores(mds1)$sites) #Using the scores function from vegan to extra
ct the site scores and convert to a
data.scores$Region = physreg$Region
data.scores$Percent = physreg$Percent
vars = as.data.frame(scores(mds1)$species)
vars$Label = row.names(vars)

ggplot(data.scores, aes(x = NMDS1, y = NMDS2))+ geom_point(aes(size = Percent), alpha = 0.5, col
or = "lightskyblue")+
  geom_text(aes(label = Region))+
  geom_text(data = vars, aes(label = Label), color = "red")+
  geom_segment(data = vars, aes(x = 0, y =0, xend = NMDS1, yend = NMDS2),
    arrow = arrow(length = unit(0.1,"cm")), inherit.aes = FALSE,
    color = "red")+
  scale_size_continuous(range = c(1,10), name = "Percent\nof samples \nwith Microcysits\nPresen
t")+
  theme_bw()
```



```
ggplot(physreg, aes(x = Region, y = Depth)) + geom_col()
```

