Exploratory Plots 20190113

Whitney Beck

January 13, 2019

## Data Prep from 1) Data Clean Script

## 1) script to clean and wrangle data  
library(tidyverse, verbose = F)

## Warning: package 'tidyverse' was built under R version 3.5.2

## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.5.1

## Warning: package 'dplyr' was built under R version 3.5.1

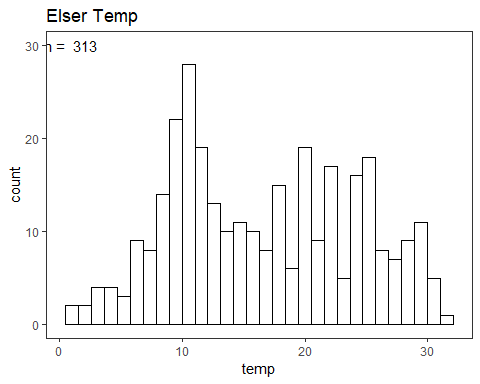
## -- Conflicts ------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2)  
theme\_mod <- function(){theme\_bw() %+replace% theme(panel.grid = element\_blank())}  
theme\_set(theme\_mod())   
  
## Data import  
Data\_Beck=read.csv("raw-data/Beck Data/Beck Data.csv")  
Data\_Elser=read.csv("raw-data/Elser-et-al-2007\_Global-N-P-limitation/data/doi\_10.5063\_AA\_nceasadmin.910.2-np-analysis-dataset-out-for-datasharing.csv")  
Data\_Vanni=read.csv("raw-data/Vanni-et-al\_2017\_Animal-excretion/Aquatic\_animal\_excretion\_data.csv")  
Data\_Vanni$Temperature..C.=as.numeric(Data\_Vanni$Temperature..C.)  
Data\_Vanni$P.excretion.rate..ug.P.ind.h.=as.numeric(Data\_Vanni$P.excretion.rate..ug.P.ind.h.)  
Data\_Vanni$N.excretion.rate..ug.N.ind.h.=as.numeric(Data\_Vanni$N.excretion.rate..ug.N.ind.h.)  
Data\_Vanni$Excreted.N.P..molar.=as.numeric(Data\_Vanni$Excreted.N.P..molar.)  
  
## Elser: set variables of interest  
var\_names\_Elser = c("id\_study","system", "strata", "habitat", "cat", "tax\_resp\_class",  
 "l.n.c","l.p.c","l.int.c","temp","n\_avail","n\_total","p\_avail",  
 "p\_total", "light", "location", "latitud","longitud","elev","duration")  
  
#Filter to variables of interest that have temperature values in aquatic systems  
TempData\_Elser <- Data\_Elser %>%  
 select(var\_names\_Elser) %>%  
 filter(system %in% c("FRESHWATER", "MARINE")) %>%  
 filter(!is.na(temp))  
  
##Beck: set variables of interest  
var\_names\_Beck=c("Ref", "Site.Name", "Latitude\_GoogleEarth", "Longitude\_GoogleEarth", "Days", "Nitrogen.Molarity",  
 "Phosphorus.Molarity", "Elevation\_m", "Depth\_m", "Canopy\_Percent", "Canopy\_Qual", "Temp\_C",  
 "NH4\_ug\_L", "NO3\_ug\_L", "DIN\_ug\_L", "TN\_mg\_L", "SRP\_ug\_L", "TP\_ug\_L", "NO3\_SRP\_Ratio",   
 "Control\_mean\_mg\_m2", "Control.SE",   
 "N\_mean\_mg\_m2", "N.SE", "N.LRR.Effect", "N.LRR.Variance",  
 "P\_mean\_mg\_m2", "P.SE", "P.LRR.Effect", "P.LRR.Variance",  
 "NP\_mean\_mg\_m2", "NP.SE", "NP.LRR", "NP.LRR.Variance")  
  
#Filter to variables of interest that have temperature values  
TempData\_Beck <- Data\_Beck %>%  
 select(var\_names\_Beck) %>%  
 filter(!is.na(Temp\_C))  
  
##Vanni: set variables of interest  
  
#Many of these could be interesting to explore, and variable names are really strangely formatted  
#so I'm going to leave as-is for now  
  
TempData\_Vanni <- Data\_Vanni%>%  
 filter(!is.na(Temperature..C.))

## Elser Plots from 2) Exploratory script

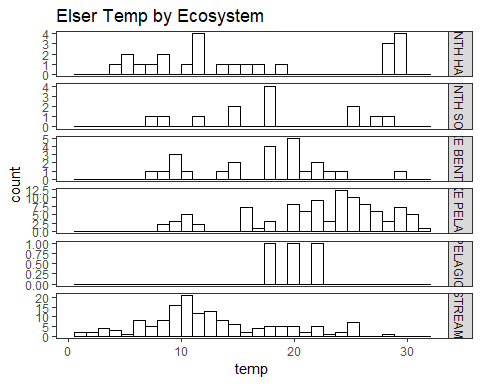
#Temperature Range  
  
ggplot(TempData\_Elser, aes(x = temp)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1, y = 30,label = paste("n = ", nrow(subset(TempData\_Elser, temp!="NA"))))+  
 ggtitle("Elser Temp")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(TempData\_Elser, aes(x = temp)) + geom\_histogram(colour = 'black',fill = 'white') + facet\_grid(cat ~ ., scales = 'free')+  
 ggtitle("Elser Temp by Ecosystem")

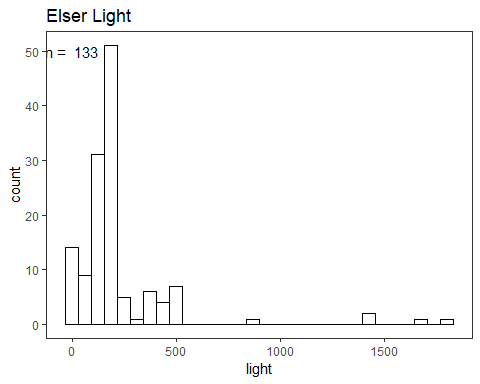
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Light Range  
  
ggplot(TempData\_Elser, aes(x = light)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1, y = 50,label = paste("n = ",nrow(subset(TempData\_Elser, light!="NA"))))+  
 ggtitle("Elser Light")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

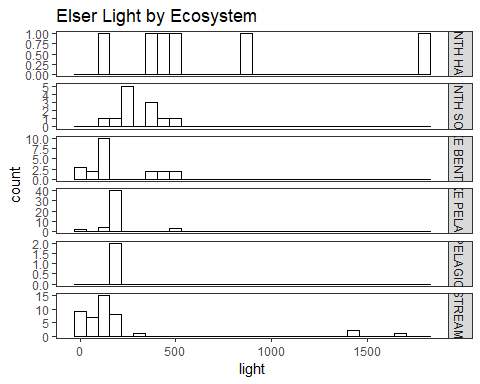
## Warning: Removed 180 rows containing non-finite values (stat\_bin).



ggplot(TempData\_Elser, aes(x = light)) + geom\_histogram(colour = 'black',fill = 'white') + facet\_grid(cat ~ ., scales = 'free')+  
 ggtitle("Elser Light by Ecosystem")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 180 rows containing non-finite values (stat\_bin).



#Avail N Range  
  
Elser\_log\_n\_avail = ggplotGrob(ggplot(TempData\_Elser, aes(x = log10(n\_avail))) + geom\_histogram(colour = 'black', fill = 'white'))

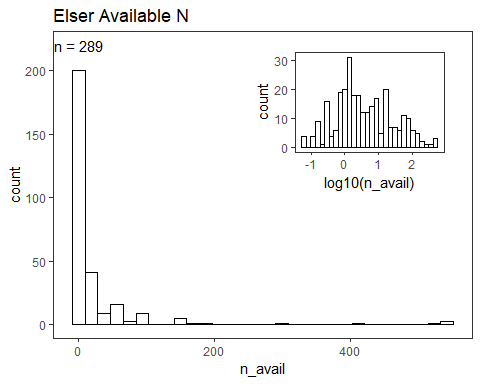
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 24 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Elser, aes(x = n\_avail)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 1, y = 220, label = paste("n =",nrow(subset(TempData\_Elser, n\_avail!="NA")))) +  
 annotation\_custom(grob = Elser\_log\_n\_avail, xmin = 250, xmax = 550, ymin = 100, ymax = 220)+  
 ggtitle("Elser Available N")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 24 rows containing non-finite values (stat\_bin).



#Total N Range  
  
Elser\_log\_n\_total = ggplotGrob(ggplot(TempData\_Elser, aes(x = log10(n\_total))) + geom\_histogram(colour = 'black', fill = 'white'))

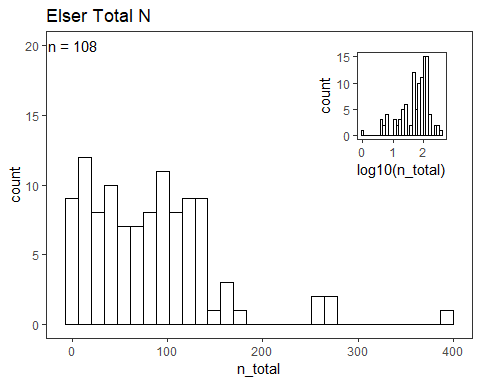
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 205 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Elser, aes(x = n\_total)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 1, y = 20, label = paste("n =", nrow(subset(TempData\_Elser, n\_total!="NA")))) +  
 annotation\_custom(grob = Elser\_log\_n\_total, xmin = 250, xmax = 400, ymin = 10, ymax = 20)+  
 ggtitle("Elser Total N")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 205 rows containing non-finite values (stat\_bin).



#Avail P Range  
  
Elser\_log\_p\_avail = ggplotGrob(ggplot(TempData\_Elser, aes(x = log10(p\_avail))) + geom\_histogram(colour = 'black', fill = 'white'))

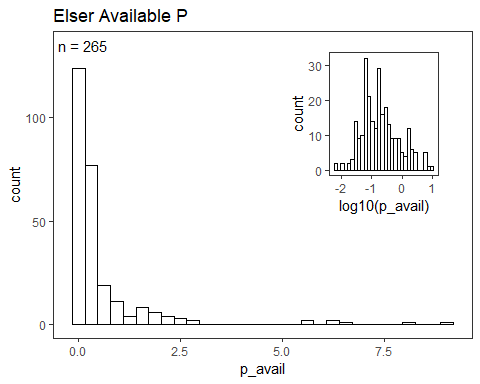
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 50 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Elser, aes(x = p\_avail)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = .1, y = 135, label = paste("n =",nrow(subset(TempData\_Elser, p\_avail!="NA")))) +  
 annotation\_custom(grob = Elser\_log\_p\_avail, xmin = 5, xmax = 9, ymin = 50, ymax = 135)+  
 ggtitle("Elser Available P")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 48 rows containing non-finite values (stat\_bin).



#Total P Range  
  
Elser\_log\_p\_total = ggplotGrob(ggplot(TempData\_Elser, aes(x = log10(p\_total))) + geom\_histogram(colour = 'black', fill = 'white'))

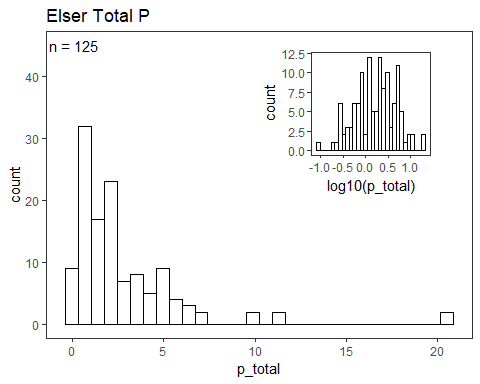
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 188 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Elser, aes(x = p\_total)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = .1, y = 45, label = paste("n =",nrow(subset(TempData\_Elser, p\_total!="NA")))) +  
 annotation\_custom(grob = Elser\_log\_p\_total, xmin = 10, xmax = 20, ymin = 20, ymax = 45)+  
 ggtitle("Elser Total P")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

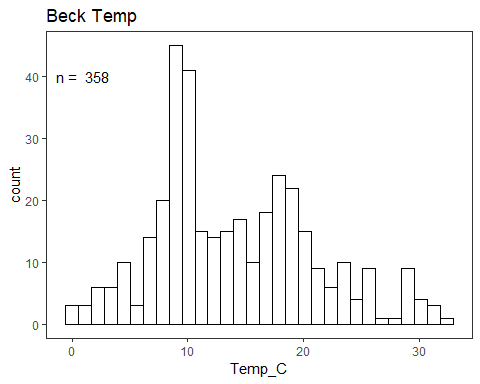
## Warning: Removed 188 rows containing non-finite values (stat\_bin).



## Beck Plots from 2) Exploratory script

##Beck Plots  
  
#Temp Range  
  
ggplot(TempData\_Beck, aes(x = Temp\_C)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1, y = 40,label = paste("n = ", nrow(subset(TempData\_Beck, Temp\_C!="NA"))))+  
 ggtitle("Beck Temp")

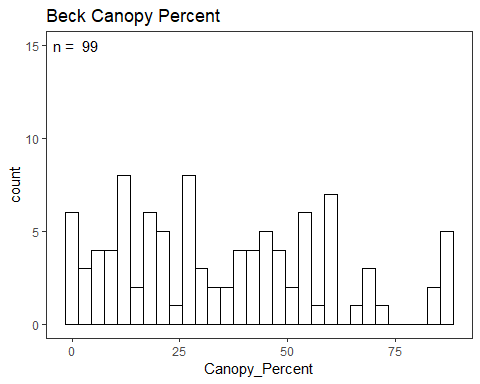
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Canopy % Range  
  
ggplot(TempData\_Beck, aes(x = Canopy\_Percent)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1, y = 15,label = paste("n = ", nrow(subset(TempData\_Beck, Canopy\_Percent!="NA"))))+  
 ggtitle("Beck Canopy Percent")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 259 rows containing non-finite values (stat\_bin).



#DIN Range  
  
Beck\_log\_DIN = ggplotGrob(ggplot(TempData\_Beck, aes(x = log10(DIN\_ug\_L))) + geom\_histogram(colour = 'black', fill = 'white'))

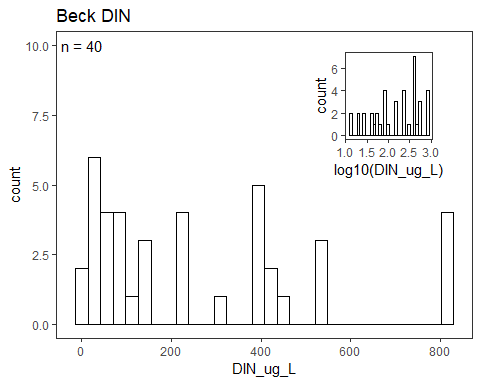
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 318 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Beck, aes(x = DIN\_ug\_L)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 1, y = 10, label = paste("n =", nrow(subset(TempData\_Beck, DIN\_ug\_L!="NA")))) +  
 annotation\_custom(grob = Beck\_log\_DIN, xmin = 500, xmax = 800, ymin = 5, ymax = 10)+  
 ggtitle("Beck DIN")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 318 rows containing non-finite values (stat\_bin).



#TN Range  
  
Beck\_log\_TN = ggplotGrob(ggplot(TempData\_Beck, aes(x = log10(TN\_mg\_L))) + geom\_histogram(colour = 'black', fill = 'white'))

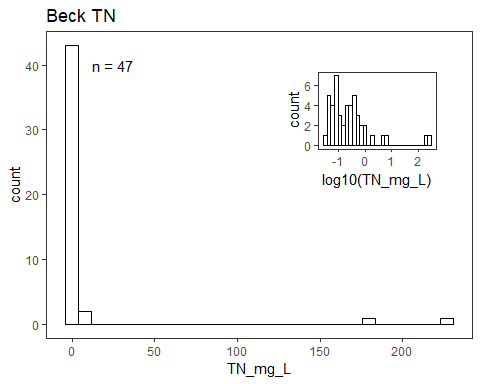
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 311 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Beck, aes(x = TN\_mg\_L)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 25, y = 40, label = paste("n =", nrow(subset(TempData\_Beck, TN\_mg\_L!="NA")))) +  
 annotation\_custom(grob = Beck\_log\_TN, xmin = 125, xmax = 225, ymin = 20, ymax = 40)+  
 ggtitle("Beck TN")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 311 rows containing non-finite values (stat\_bin).



#SRP Range  
  
Beck\_log\_SRP = ggplotGrob(ggplot(TempData\_Beck, aes(x = log10(SRP\_ug\_L))) + geom\_histogram(colour = 'black', fill = 'white'))

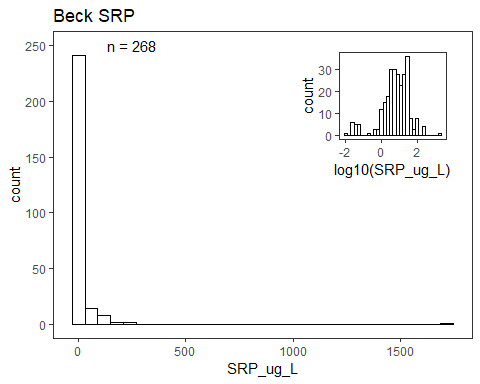
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 90 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Beck, aes(x = SRP\_ug\_L)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 250, y = 250, label = paste("n =", nrow(subset(TempData\_Beck, SRP\_ug\_L!="NA")))) +  
 annotation\_custom(grob = Beck\_log\_SRP, xmin = 1000, xmax = 1750, ymin = 125, ymax = 250)+  
 ggtitle("Beck SRP")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 90 rows containing non-finite values (stat\_bin).



#TP Range  
  
Beck\_log\_TP = ggplotGrob(ggplot(TempData\_Beck, aes(x = log10(TP\_ug\_L))) + geom\_histogram(colour = 'black', fill = 'white'))

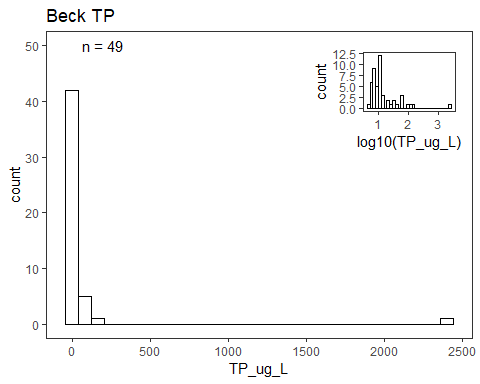
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 309 rows containing non-finite values (stat\_bin).

ggplot(TempData\_Beck, aes(x = TP\_ug\_L)) + geom\_histogram(colour = 'black', fill = 'white') +  
 annotate('text', x = 200, y = 50, label = paste("n =", nrow(subset(TempData\_Beck, TP\_ug\_L!="NA")))) +  
 annotation\_custom(grob = Beck\_log\_TP, xmin = 1500, xmax = 2500, ymin = 30, ymax = 50)+  
 ggtitle("Beck TP")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

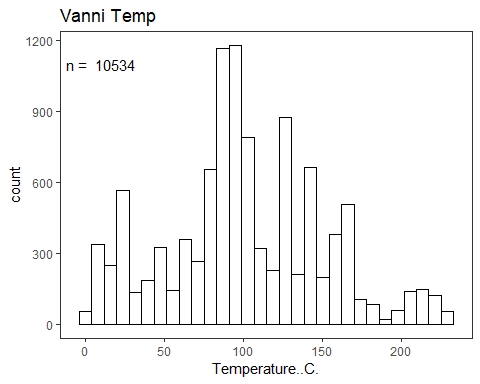
## Warning: Removed 309 rows containing non-finite values (stat\_bin).



## Vanni Plots from 2) Exploratory script

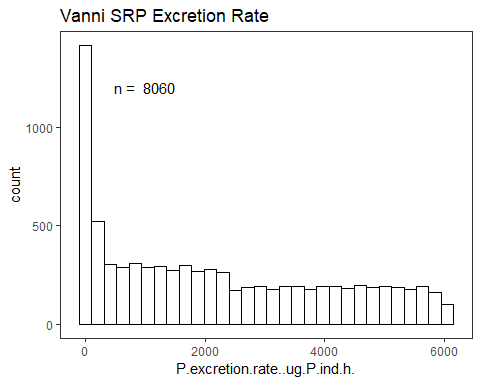
#Temp Range  
  
ggplot(TempData\_Vanni, aes(x = Temperature..C.)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 10, y = 1100,label = paste("n = ", nrow(subset(TempData\_Vanni, Temperature..C.!="NA"))))+  
 ggtitle("Vanni Temp")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



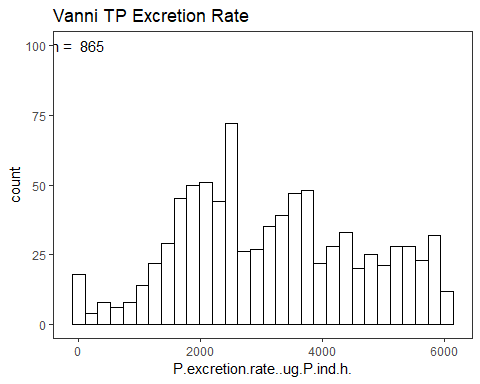
#Excretion SRP  
  
ggplot(subset(TempData\_Vanni, P.form=="SRP"), aes(x = P.excretion.rate..ug.P.ind.h.)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1000, y = 1200,label = paste("n = ", nrow(subset(TempData\_Vanni, P.form=="SRP"& P.excretion.rate..ug.P.ind.h.!="NA"))))+  
 ggtitle("Vanni SRP Excretion Rate")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



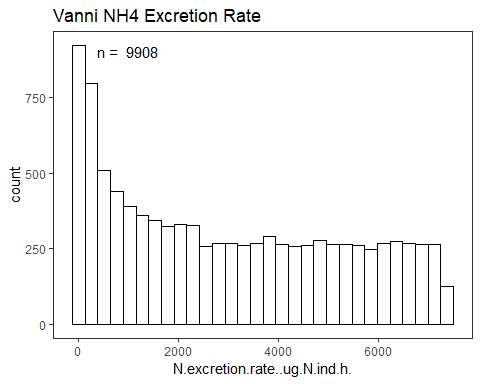
#Excretion TP  
  
ggplot(subset(TempData\_Vanni, P.form=="TP"), aes(x = P.excretion.rate..ug.P.ind.h.)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 0, y = 100,label = paste("n = ", nrow(subset(TempData\_Vanni, P.form=="TP"& P.excretion.rate..ug.P.ind.h.!="NA"))))+  
 ggtitle("Vanni TP Excretion Rate")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



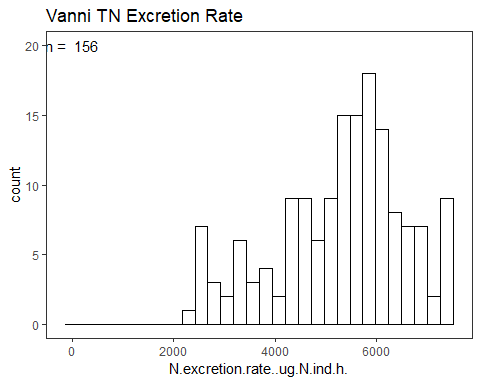
#Excretion Ammonium  
  
ggplot(subset(TempData\_Vanni, N.form=="NH4"), aes(x = N.excretion.rate..ug.N.ind.h.)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 1000, y = 900,label = paste("n = ", nrow(subset(TempData\_Vanni, N.form=="NH4" &N.excretion.rate..ug.N.ind.h.!="NA"))))+  
 ggtitle("Vanni NH4 Excretion Rate")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Excretion TN  
  
ggplot(subset(TempData\_Vanni, N.form=="TN"), aes(x = N.excretion.rate..ug.N.ind.h.)) + geom\_histogram(colour = 'black', fill = 'white')+  
 annotate('text', x = 0, y = 20,label = paste("n = ", nrow(subset(TempData\_Vanni, N.form=="TN" &N.excretion.rate..ug.N.ind.h.!="NA"))))+  
 ggtitle("Vanni TN Excretion Rate")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Excretion N:P (If N excretion or P excretion was NA, value was calculated as "1",   
#so only included values >1)  
  
ggplot(subset(TempData\_Vanni, Excreted.N.P..molar.>1), aes(x = Excreted.N.P..molar.)) + geom\_histogram(colour = 'black', fill = 'white', bins=50)+  
 annotate('text', x = 500, y = 300,label = paste("n = ", nrow(subset(TempData\_Vanni, Excreted.N.P..molar.!="NA" & Excreted.N.P..molar.>1))))+  
 ggtitle("Vanni Excretion N:P Molar")

