MICB 405 Project 2

Shannah

November 17, 2018

First I will analyse the MAGs from checkM output

library(tidyverse)  
  
#Load .tsv file from MetaBAT2 to dat  
bin\_dat <- read\_tsv(file="MetaBAT2\_SaanichInlet\_200m\_min1500\_checkM\_stdout.tsv", col\_names = TRUE)  
  
#Rename Marker lineage so it's one string (makes it easier later)  
bin\_dat <- bin\_dat %>%  
 dplyr::rename('Marker\_lineage' = 'Marker lineage')  
  
#Add new column with just lineage for point colors  
bin\_dat <- bin\_dat %>%  
 dplyr::mutate('lineage' = gsub(".\*\_\_(\*)", "\\1", Marker\_lineage),  
 'lineage' = gsub("\\(|\\)", "", lineage),  
 'lineage' = gsub("UID.\*", "", lineage))  
  
#Plotting Completeness vs Contamination of MAGs at 200m  
bin\_dat %>%  
 ggplot(aes(x=Completeness, y=Contamination, colour=lineage)) +  
 geom\_point() +  
 scale\_color\_manual(values = c('#a6cee3','#1f78b4',"#b2df8a","#33a02c","#fb9a99",  
 "#e31a1c",'#fdbf6f',"#ff7f00","#cab2d6","#6a3d9a",  
 "#ffff99","#b15928", "#969696")) #added discrete scales ; palette from colorbrewer2 but added extra ("#969696") b/c 13 values

