7659\_HW4

Guannan Shen

October 15, 2018

Table of Contents

## set up workspace  
library(knitr)  
library(tidyverse)  
options(stringsAsFactors = F)  
options(dplyr.width = Inf)  
getwd()

## [1] "/home/guanshim/Documents/Stats/CIDA\_OMICs/7659Stats\_Genetics/HW4"

gal\_sum <- read.table("Galaxy3\_FASTQ\_Summary\_Statistics\_on\_data\_2.tabular")  
ggplot(gal\_sum, aes(x = V1)) + geom\_line(aes(y = V14, color = "A\_Count")) +   
 geom\_point(aes(y = V14, color = "A\_Count")) + geom\_line(aes(y = V15,   
 color = "C\_Count")) + geom\_point(aes(y = V15, color = "C\_Count")) +   
 geom\_line(aes(y = V16, color = "G\_Count")) + geom\_point(aes(y = V16,   
 color = "G\_Count")) + geom\_line(aes(y = V17, color = "T\_Count")) +   
 geom\_point(aes(y = V17, color = "T\_Count")) + geom\_line(aes(y = V18,   
 color = "N\_Count")) + geom\_point(aes(y = V18, color = "N\_Count")) +   
 scale\_x\_discrete(name = "Read Positions", limits = c(1:36)) +   
 scale\_y\_continuous(name = "Nucleotide Content") + theme\_bw() +   
 scale\_colour\_manual("", breaks = c("A\_Count", "C\_Count",   
 "G\_Count", "T\_Count", "N\_Count"), values = c(A\_Count = "red",   
 C\_Count = "black", G\_Count = "grey", T\_Count = "yellow",   
 N\_Count = "brown"))

