# installing & loading package---------

library(tidyverse)

library(here)

library(janitor)

library(skimr)

# now read data in data folder-----

beaches <- read\_csv(here("data","sydneybeaches.csv"))

# exploring the data-------

View(beaches)

dim(beaches)

head(beaches)

tail(beaches)

glimpse(beaches)

summary(beaches)

skim(beaches)

# tidying columns--------

select\_all(beaches,toupper)

select\_all(beaches,tolower)

cleanbeaches <- clean\_names(beaches)

cleanbeaches <- read\_csv(here("data","sydneybeaches.csv"))

cleanbeaches <- select\_all(cleanbeaches,tolower)

# selecting few variable-----

select(cleanbeaches,site,council,beachbugs,everything())

# which beach has a extreme level of bugs-------

worstbeach <- cleanbeaches %>%

arrange(desc(beachbugs))

# using filter and arrange for coogee beach-------

worstcoogee <- cleanbeaches %>%

filter(site == "Coogee Beach") %>%

arrange(desc(beachbugs))

# coogee and bondi statistic-------

coogee\_bondi <- cleanbeaches %>%

filter(site == "Coogee Beach" | site == "Bondi Beach") %>%

group\_by(site) %>%

summarise(minbug = min(beachbugs,na.rm = TRUE),

maxbug = max(beachbugs,na.rm = TRUE),

meanbug = mean(beachbugs,na.rm = TRUE),

sdbug = sd(beachbugs,na.rm = TRUE))

# statistical information of all site-------

all\_beaches <- cleanbeaches %>%

group\_by(site) %>%

summarise(minbug = min(beachbugs,na.rm = TRUE),

maxbug = max(beachbugs,na.rm = TRUE),

meanbug = mean(beachbugs,na.rm = TRUE),

sdbug = sd(beachbugs,na.rm = TRUE))

# lets compare council----------

councilbysite <- cleanbeaches %>%

group\_by(council,site) %>%

summarise(meanbug = mean(beachbugs,na.rm = TRUE),

medianbug = median(beachbugs,na.rm = TRUE))

# let's create new variable(using a separate and unite function)-----

glimpse(cleanbeaches)

testdate <- cleanbeaches %>%

separate(date,c('day','month','year'),remove = FALSE)

unite\_colums <- cleanbeaches %>%

unite(council\_site,council:site,remove = FALSE)

# use mutate to transform beachbugs data-----

cleanbeaches %>% mutate(logbeachbugs = log(beachbugs))

# use mutate to compute new numeric data------

cleanbeaches %>% mutate(beachbugsdiff = beachbugs - lag(beachbugs))

# use mutate to new logical variable--------

cleanbeaches %>% mutate(buggier = beachbugs > mean(beachbugs,na.rm = TRUE))

# grouping all above function together-------

cleanbeaches\_new <- cleanbeaches %>%

separate(date,c('day','month','year'),remove = FALSE) %>%

unite(council\_site,council:site,remove = FALSE) %>%

mutate(logbeachbugs = log(beachbugs)) %>%

mutate(beachbugsdiff = beachbugs - lag(beachbugs)) %>%

mutate(buggier = beachbugs > mean(beachbugs,na.rm = TRUE))

# baker data for pivot\_longer & pivot\_wider--------

baker <- read\_csv(here("data","bakers.csv"))

baker\_long <- baker %>%

pivot\_longer(names\_to = 'spices',

values\_to = 'correct',

cinnamon\_1:nutmeg\_3)

# now beachbug data for pivot longer-------

beachbug\_long <- read\_csv(here("data","beachbugs\_wide.csv"))

new\_beachbug\_long <- beachbug\_long %>%

pivot\_longer(names\_to = 'site',

values\_to = 'beachbugs',

cols = -year)

# back to wide data format--------

new\_beachbug\_wide <- new\_beachbug\_long %>%

pivot\_wider(names\_from = site,

values\_from = beachbugs)

# new frame wide data------

frame\_wide <- read\_csv(here("data","frames\_wide.csv"))

new\_frame\_long <- frame\_wide %>%

pivot\_longer(names\_to = 'size','item',

values\_to = 'reponse',

cols = large\_item1:small\_item7)

# write back csv file to data folder-----

write\_csv(cleanbeaches\_new,here("data","cleanbeaches\_new.csv"))