Irene Data analysis

Mwaura Patrick

1/27/2021

options( warn = -1 )  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.5 v dplyr 1.0.3  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

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

library(dplyr)  
library(readxl)  
library(here)

## here() starts at E:/RESULTS/Mwaura/mwaura/WORK1/Covid19

library(skimr)  
library(kableExtra)

##   
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':  
##   
## group\_rows

library(AMR)  
library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

library(pipeR)  
library(cleaner)

## Set the working directory and load the data

setwd("E:/RESULTS/Mwaura/mwaura/WORK1/Covid19/Irene Work")  
IRENE\_DATA\_ANALYSIS\_R <- read\_csv("IRENE DATA ANALYSIS R.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## .default = col\_character(),  
## CAMEL\_ID\_NUMBER = col\_double(),  
## MDR = col\_double()  
## )  
## i Use `spec()` for the full column specifications.

## explore the data

#skimr::skim(IRENE\_DATA\_ANALYSIS\_R)

## Brief summary

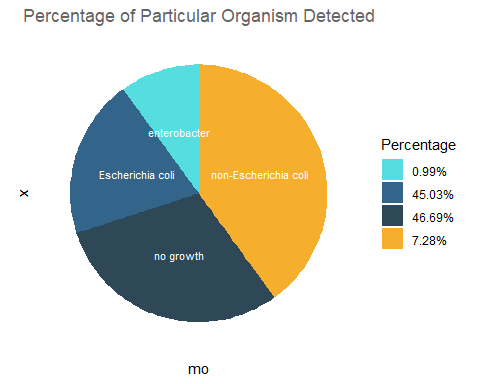
IRENE\_DATA\_ANALYSIS\_R %>% select(mo,CAMEL\_ID\_NUMBER) %>% group\_by(mo) %>% summarize(Count=n()) %>% mutate(Percentage=paste0(round(Count/sum(Count)\*100,2),"%"))

## # A tibble: 4 x 3  
## mo Count Percentage  
## \* <chr> <int> <chr>   
## 1 enterobacter 3 0.99%   
## 2 Escherichia coli 136 45.03%   
## 3 no growth 141 46.69%   
## 4 non-Escherichia coli 22 7.28%

IRENE\_DATA\_ANALYSIS\_R %>% freq(mo)

##   
##   
## \*\*Frequency table\*\*   
##   
## Class: character   
## Length: 302   
## Available: 302 (100%, NA: 0 = 0%)   
## Unique: 4   
##   
## Shortest: 9   
## Longest: 20  
##   
##   
## | |Item | Count| Percent| Cum. Count| Cum. Percent|  
## |:--|:--------------------|-----:|-------:|----------:|------------:|  
## |1 |no growth | 141| 46.69%| 141| 46.69%|  
## |2 |Escherichia coli | 136| 45.03%| 277| 91.72%|  
## |3 |non-Escherichia coli | 22| 7.28%| 299| 99.01%|  
## |4 |enterobacter | 3| 0.99%| 302| 100.00%|

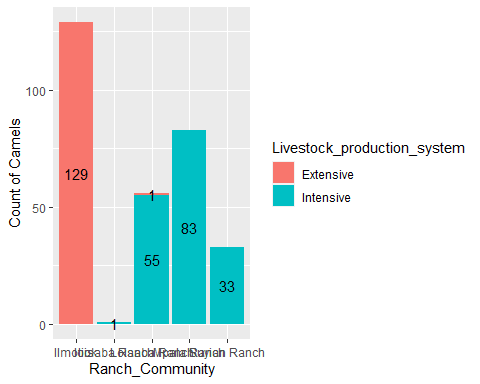
IRENE\_DATA\_ANALYSIS\_R %>% select(mo,CAMEL\_ID\_NUMBER) %>% group\_by(mo) %>% summarize(Count=n()) %>% mutate(Percentage=paste0(round(Count/sum(Count)\*100,2),"%")) %>% ggplot(aes(x="", y=mo, fill=Percentage)) + geom\_bar(stat="identity", width=1)+ coord\_polar("y", start=0) +theme\_classic()+labs( title = "Percentage of Particular Organism Detected")+scale\_fill\_manual(values=c("#55DDE0", "#33658A", "#2F4858", "#F6AE2D"))+theme(axis.line = element\_blank(),axis.text = element\_blank(),axis.ticks = element\_blank(), plot.title = element\_text(hjust = 0.5, color = "#666666"))+geom\_text(aes(y = mo, label = mo), position = position\_stack(vjust = .5),color = "white", size=3)

 ## cross tabulation by Ranch\_community vs the livestock production system

IRENE\_DATA\_ANALYSIS\_R %>% select(Ranch\_Community,Livestock\_production\_system) %>% pivot\_wider(names\_from = Livestock\_production\_system, values\_from= Livestock\_production\_system,values\_fn=list(Livestock\_production\_system=list))

## # A tibble: 5 x 3  
## Ranch\_Community Intensive Extensive   
## <chr> <list> <list>   
## 1 Mpala Ranch <chr [83]> <NULL>   
## 2 Ilmotiok <NULL> <chr [129]>  
## 3 Loisaba Ranch <chr [55]> <chr [1]>   
## 4 loisaba Ranch <chr [1]> <NULL>   
## 5 Suyian Ranch <chr [33]> <NULL>

IRENE\_DATA\_ANALYSIS\_R %>% select(Ranch\_Community,Livestock\_production\_system) %>% ggplot(aes(x=Ranch\_Community,fill=Livestock\_production\_system))+geom\_bar()+ylab("Count of Camels")+stat\_count(geom="text",aes(label=stat(count)), position = position\_stack(vjust = .5),color="black")

 ## MDRO analysis

IRENE\_DATA\_ANALYSIS\_R %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin) %>% mutate\_at(vars(Amoxicillin\_clavulanic\_acid:Norfloxacin), as.rsi)%>% mdro()%>% freq()

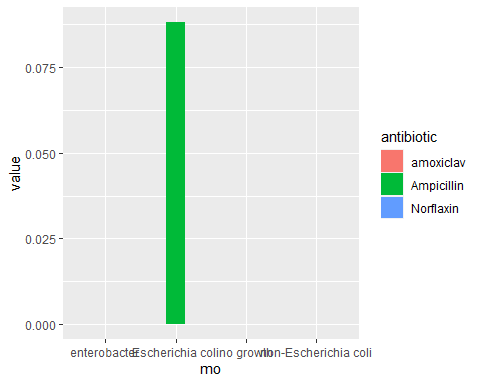
##   
##   
## \*\*Frequency table\*\*   
##   
## Class: factor > ordered (numeric)   
## Length: 302   
## Levels: 4: Negative < Multi-drug-resistant (MDR) < Extensively drug-resistant ...   
## Available: 158 (52.32%, NA: 144 = 47.68%)   
## Unique: 2  
##   
##   
## | |Item | Count| Percent| Cum. Count| Cum. Percent|  
## |:--|:--------------------------|-----:|-------:|----------:|------------:|  
## |1 |Negative | 152| 96.20%| 152| 96.20%|  
## |2 |Multi-drug-resistant (MDR) | 6| 3.80%| 158| 100.00%|

## Resistance

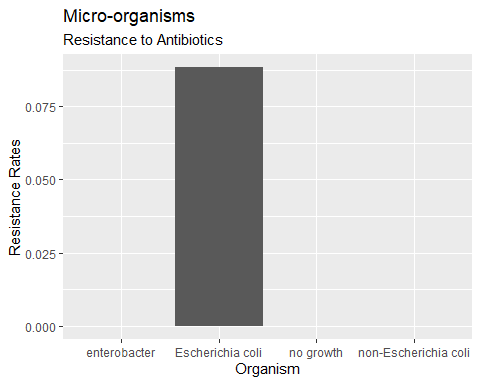
IRENE\_DATA\_ANALYSIS\_R %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin) %>% mutate\_at(vars(Amoxicillin\_clavulanic\_acid:Norfloxacin), as.rsi) %>% group\_by(mo) %>% summarise(amoxiclav = resistance(Amoxicillin\_clavulanic\_acid), Ampicillin = resistance(Ampicillin), Norflaxin = resistance(Norfloxacin))

## # A tibble: 4 x 4  
## mo amoxiclav Ampicillin Norflaxin  
## \* <chr> <dbl> <dbl> <dbl>  
## 1 enterobacter NA NA NA  
## 2 Escherichia coli 0 0.0882 0  
## 3 no growth NA NA NA  
## 4 non-Escherichia coli NA NA NA

IRENE\_DATA\_ANALYSIS\_R %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin) %>% mutate\_at(vars(Amoxicillin\_clavulanic\_acid:Norfloxacin), as.rsi) %>% group\_by(mo) %>% summarise(amoxiclav = resistance(Amoxicillin\_clavulanic\_acid), Ampicillin = resistance(Ampicillin), Norflaxin = resistance(Norfloxacin)) %>% tidyr::pivot\_longer(-mo, names\_to = "antibiotic") %>% ggplot(aes(x = mo,y = value,fill = antibiotic)) +geom\_col(position = "dodge2")



IRENE\_DATA\_ANALYSIS\_R %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin) %>% mutate\_at(vars(Amoxicillin\_clavulanic\_acid:Norfloxacin), as.rsi) %>% group\_by(mo) %>% summarise(amoxiclav = resistance(Amoxicillin\_clavulanic\_acid), Ampicillin = resistance(Ampicillin), Norflaxin = resistance(Norfloxacin)) %>% tidyr::pivot\_longer(-mo, names\_to = "antibiotic") %>%ggplot(mapping = aes(x = mo,y = value)) +geom\_col() +labs(title = "Micro-organisms",subtitle = "Resistance to Antibiotics", x = "Organism",y = "Resistance Rates")

 ## Conduct principal component analysis (PCA) for AMR

resistance\_data<-IRENE\_DATA\_ANALYSIS\_R %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin) %>% mutate\_at(vars(Amoxicillin\_clavulanic\_acid:Norfloxacin), as.rsi) %>% group\_by(mo) %>% summarise\_if(is.rsi, resistance) %>% select(mo,Amoxicillin\_clavulanic\_acid:Norfloxacin)  
  
resistance\_data

## # A tibble: 4 x 16  
## mo Amoxicillin\_cla~ Ampicillin Ceftazidime Cefotaxime Ceftriaxone  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 ente~ NA NA NA NA NA   
## 2 Esch~ 0 0.0882 0.0735 0.162 0.0294  
## 3 no g~ NA NA NA NA NA   
## 4 non-~ NA NA NA NA NA   
## # ... with 10 more variables: Cefuroxime <dbl>, Cefepime <dbl>, Cefaclor <dbl>,  
## # Tetracycline <dbl>, Gentamicin <dbl>, Spectinomycin <dbl>,  
## # Chloramphenicol <dbl>, Ciprofloxacin <dbl>,  
## # Trimethoprim\_sulfamethoxazole <dbl>, Norfloxacin <dbl>

resistance\_data[resistance\_data==0]<-NA  
  
  
resistant\_data\_2<-data.frame(na.omit(resistance\_data))  
  
resistant\_data\_2

## [1] mo Amoxicillin\_clavulanic\_acid   
## [3] Ampicillin Ceftazidime   
## [5] Cefotaxime Ceftriaxone   
## [7] Cefuroxime Cefepime   
## [9] Cefaclor Tetracycline   
## [11] Gentamicin Spectinomycin   
## [13] Chloramphenicol Ciprofloxacin   
## [15] Trimethoprim\_sulfamethoxazole Norfloxacin   
## <0 rows> (or 0-length row.names)