

```

library(data.table)
library(lubridate)
library(ggplot2)
rm(list = ls())

setwd('FILEPATH')

master_data1 <- fread('FILEPATH')
master_data1

mydata1 <- master_data1[,.(SampleID, profile, Organism,
                           Ampicillin, Cotrimoxazole, Chloramphenicol,
                           Penicillin, Gentamicin, Ceftriaxone,
                           Ciprofloxacin, Ceftizadime,
                           Gender, Age, Year)]

setnames(mydata1,
         old = c("SampleID", "profile", "Organism",
                  'Gender', 'Age', 'Year'),
         new = c("specimen_id", "specimen_source", "org_name",
                  "sex", "age", "year"))

```



```

mydata1 <- melt(mydata1,
               id.vars = c("specimen_id", "specimen_source",
                           "org_name", "sex", "age",
                           "year"),
               variable.name = "drug_name",
               value.name = "interpretation")

```

```

mydata1$specimen_source[mydata1$specimen_source == 'BC'] <- 'blood'

```

```

mydata1 <- mydata1[order(mydata1$specimen_id),]
mydata1

```

```

mydata1$cephalosporins <- 0

```



```

mydata1$cephalosporins[grepl('CEFDINIR', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFIXIME', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFOTAXIME', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFTIZOXIME', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFPODOXIME', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFTAZIDIME', mydata1$drug_name, ignore.case = T)|
                       grepl('Ceftizadime', mydata1$drug_name, ignore.case = T)|
                       grepl('CEFTRIAXONE', mydata1$drug_name, ignore.case = T)|
                       grepl('Ceftriaxone', mydata1$drug_name, ignore.case = T)] <-

```

1

```

mydata1$penicillins <- 0

```

```

mydata1$penicillins[grepl('PENICILLIN', mydata1$drug_name, ignore.case = T)] <- 1

```

```

mydata1$penicillins[grepl('Ampicillin', mydata1$drug_name, ignore.case = T)] <- 1

```

```

mydata1$penicillins[grepl('OXACILLIN', mydata1$drug_name, ignore.case = T)] <- 1

```

```

mydata1$MRSA_test <- 0

```

```

mydata1$MRSA_test[grepl('METHICILLIN', mydata1$drug_name, ignore.case = T)] <- 1

```

```

      grepl('OXACILLIN', mydata1$drug_name, ignore.case = T) |
      grepl('CEFOXITIN', mydata1$drug_name, ignore.case = T)] <- 1

sort(unique(mydata1$org_name))
mydata1 <- mydata1[(grepl('S. aureus', mydata1$org_name, ignore.case = T) &
mydata1$MRSA_test == 1)|
      (grepl('MRSA', mydata1$org_name, ignore.case = T) &
mydata1$MRSA_test == 1)|
      (grepl('S. pneumoniae', mydata1$org_name, ignore.case = T) &
mydata1$penicillins == 1) |
      (grepl('escherichia coli', mydata1$org_name, ignore.case = T) &
mydata1$cephalosporins == 1)|
      (grepl('klebsiella pneumoniae', mydata1$org_name, ignore.case = T) &
mydata1$cephalosporins == 1) |
      (grepl('Klebsiella pneumoniae', mydata1$org_name, ignore.case = T) &
mydata1$cephalosporins == 1),]

mydata1$org_name[grepl('S. aureus', mydata1$org_name, ignore.case = T)] <-
'Staphylococcus aureus'
mydata1$org_name[grepl('MRSA', mydata1$org_name, ignore.case = T)] <- 'Staphylococcus
aureus'
mydata1$org_name[grepl('S. pneumoniae', mydata1$org_name, ignore.case = T)] <-
'Streptococcus pneumoniae'
mydata1$org_name[grepl('escherichia coli', mydata1$org_name, ignore.case = T)] <-
'Escherichia coli'
mydata1$org_name[grepl('klebsiella pneumoniae', mydata1$org_name, ignore.case = T)] <-
'Klebsiella pneumoniae'

mydata1$resistance <- NA
mydata1$resistance[mydata1$interpretation == 'R'] <- 1
mydata1$resistance[mydata1$interpretation == 'S'] <- 0
mydata1$resistance[mydata1$interpretation == 'Intermediate'] <- 1
mydata1$resistance[mydata1$interpretation == 'Positive'] <- 1
mydata1$resistance[mydata1$interpretation == 'Non Suceptible'] <- 1
mydata1$resistance[mydata1$interpretation == 'Negative'] <- 0

mydata1 <- mydata1[!is.na(mydata1$resistance),]
mydata1

master_data2 <- fread('FILEPATH')

mydata2 <- master_data2[,.(RequestNumber, ProfileName, organism,
      Ampicillin10, Penicillin1, CHLO,
      Cotrimoxazole25, CEFTR_Cefpodoxime, Ciprofloxacin1,
      Gentamicin10, Ceftriaxone30,
      Cefoxitin10_Clox, Ceftazidime, Erythromycin5,
      Oxacillin1_Pen, Tetracycline10, Amikacin30,
      Augmentin30, Gender, PatientAge, YEAR)]

setnames(mydata2,
      old = c("RequestNumber", "ProfileName", "organism",
      'Ampicillin10', 'Penicillin1', 'CHLO',
      'Cotrimoxazole25', 'CEFTR_Cefpodoxime', 'Ciprofloxacin1',
      'Gentamicin10', 'Ceftriaxone30',
      'Cefoxitin10_Clox', 'Ceftazidime', 'Erythromycin5',
      'Oxacillin1_Pen', 'Tetracycline10', 'Amikacin30',

```

```

      'Augmentin30', 'Gender', 'PatientAge', 'YEAR'),
new = c("specimen_id", "specimen_source", "org_name",
      'Ampicillin', 'Penicillin', 'Chloramphenicol',
      'Cotrimoxazole', 'Cefpodoxime', 'Ciprofloxacin',
      'Gentamicin', 'Ceftriaxone', 'Cefoxitin',
      'Ceftazidime', 'Erythromycin', 'Oxacillin', 'Tetracycline',
      'Amikacin', 'Augmentin', "sex", "age", "year"))

```

```

mydata2$Ampicillin[mydata2$Ampicillin==1] <- 'R'
mydata2$Ampicillin[mydata2$Ampicillin==0] <- 'S'
mydata2$Penicillin[mydata2$Penicillin==1] <- 'R'
mydata2$Penicillin[mydata2$Penicillin==0] <- 'S'
mydata2$Chloramphenicol[mydata2$Chloramphenicol==1] <- 'R'
mydata2$Chloramphenicol[mydata2$Chloramphenicol==0] <- 'S'
mydata2$Cotrimoxazole[mydata2$Cotrimoxazole==1] <- 'R'
mydata2$Cotrimoxazole[mydata2$Cotrimoxazole==0] <- 'S'
mydata2$Cefpodoxime[mydata2$Cefpodoxime==1] <- 'R'
mydata2$Cefpodoxime[mydata2$Cefpodoxime==0] <- 'S'
mydata2$Gentamicin[mydata2$Gentamicin==1] <- 'R'
mydata2$Gentamicin[mydata2$Gentamicin==0] <- 'S'
mydata2$Ceftriaxone[mydata2$Ceftriaxone==1] <- 'R'
mydata2$Ceftriaxone[mydata2$Ceftriaxone==0] <- 'S'
mydata2$Cefoxitin[mydata2$Cefoxitin==1] <- 'R'
mydata2$Cefoxitin[mydata2$Cefoxitin==0] <- 'S'
mydata2$Ceftazidime[mydata2$Ceftazidime==1] <- 'R'
mydata2$Ceftazidime[mydata2$Ceftazidime==0] <- 'S'
mydata2$Erythromycin[mydata2$Erythromycin==1] <- 'R'
mydata2$Erythromycin[mydata2$Erythromycin==0] <- 'S'
mydata2$Oxacillin[mydata2$Oxacillin==1] <- 'R'
mydata2$Oxacillin[mydata2$Oxacillin==0] <- 'S'
mydata2$Tetracycline[mydata2$Tetracycline==1] <- 'R'
mydata2$Tetracycline[mydata2$Tetracycline==0] <- 'S'

```

```

mydata2 <- melt(mydata2,
  id.vars = c("specimen_id", "specimen_source",
    "org_name", "sex", "age", "year"),
  variable.name = "drug_name",
  value.name = "interpretation")

```

```

mydata2$specimen_source[mydata2$specimen_source == 'Blood Culture MC&S Procedure'] <-
'blood'
mydata2$specimen_source[mydata2$specimen_source == 'BC Paediatric'] <- 'blood'
mydata2$specimen_source[mydata2$specimen_source == 'BC Adult'] <- 'blood'

```

```

mydata2$cephalosporins <- 0
mydata2$cephalosporins[grepl('CEFDINIR', mydata2$drug_name, ignore.case = T)|
  grepl('CEFIXIME', mydata2$drug_name, ignore.case = T)|
  grepl('CEFOTAXIME', mydata2$drug_name, ignore.case = T)|
  grepl('CEFTIZOXIME', mydata2$drug_name, ignore.case = T)|
  grepl('CEFPDOXIME', mydata2$drug_name, ignore.case = T)|
  grepl('CEFTAZIDIME', mydata2$drug_name, ignore.case = T)|
  grepl('Ceftizadime', mydata2$drug_name, ignore.case = T)|
  grepl('CEFTRIAXONE', mydata2$drug_name, ignore.case = T)|
  grepl('Ceftriaxiacoxone', mydata2$drug_name, ignore.case = T)] <-

```

1

```
mydata2$penicillins <- 0
mydata2$penicillins[grepl('PENICILLIN', mydata2$drug_name, ignore.case = T)] <- 1

mydata2$penicillins[grepl('Ampicillin', mydata2$drug_name, ignore.case = T)] <- 1

mydata2$penicillins[grepl('OXACILLIN', mydata2$drug_name, ignore.case = T)] <- 1

mydata2$MRSA_test <- 0
mydata2$MRSA_test[grepl('METHICILLIN', mydata2$drug_name, ignore.case = T) |
                    grepl('OXACILLIN', mydata2$drug_name, ignore.case = T) |
                    grepl('CEFOXITIN', mydata2$drug_name, ignore.case = T)] <- 1

sort(unique(mydata2$org_name))
mydata2 <- mydata2[(grepl('Methicillin Resistant Staph Aureus', mydata2$org_name,
ignore.case = T) & mydata2$MRSA_test == 1) |
                  (grepl('Staphylococcus aureus', mydata2$org_name, ignore.case = T) &
mydata2$MRSA_test == 1) |
                  (grepl('Streptococcus pneumoniae', mydata2$org_name, ignore.case =
T) & mydata2$penicillins == 1) |
                  (grepl('Escherichia coli', mydata2$org_name, ignore.case = T) &
mydata2$cephalosporins == 1) |
                  (grepl('Klebsiella pneumoniae', mydata2$org_name, ignore.case = T) &
mydata2$cephalosporins == 1),]

mydata2$org_name[grepl('Methicillin Resistant Staph Aureus', mydata2$org_name,
ignore.case = T)] <- 'Staphylococcus aureus'
mydata2$org_name[grepl('Staphylococcus aureus', mydata2$org_name, ignore.case = T)] <-
'Staphylococcus aureus'
mydata2$org_name[grepl('Streptococcus pneumoniae', mydata2$org_name, ignore.case = T)]
<- 'Streptococcus pneumoniae'
mydata2$org_name[grepl('Escherichia coli', mydata2$org_name, ignore.case = T)] <-
'Escherichia coli'
mydata2$org_name[grepl('Klebsiella pneumoniae', mydata2$org_name, ignore.case = T)] <-
'Klebsiella pneumoniae'

mydata2$resistance <- NA
mydata2$resistance[mydata2$interpretation == 'R'] <- 1
mydata2$resistance[mydata2$interpretation == 'S'] <- 0
mydata2$resistance[mydata2$interpretation == 'I'] <- 1

mydata2 <- mydata2[!is.na(mydata2$resistance),]
mydata2

mydata <- rbind(mydata1, mydata2)
mydata$year_id <- mydata$year
mydata$year <- NULL

collapsed <- mydata[,.(resistance = max(resistance)),
                     by = c("specimen_id",
                           "specimen_source",
                           "org_name",
                           "year_id",
```

```
"sex",  
'age']]
```

```
unique(collapsed$specimen_source)  
collapsed$specimen_source[collapsed$specimen_source == ""] <- 'blood'  
  
collapsed$specimen_source <- factor(collapsed$specimen_source,  
                                  levels = c('blood', "intraabdominal", "resp",  
"urine", "skin", "other"))  
  
collapsed <- collapsed[order(year_id),]  
  
collapsed$dup <- duplicated(collapsed[,.(specimen_id, org_name, year_id, age)])  
table(collapsed$dup)  
  
dup_ids <- collapsed$specimen_id[collapsed$dup == TRUE]  
duped <- collapsed[collapsed$specimen_id %in% dup_ids,]  
duped  
  
collapsed <- collapsed[collapsed$dup == FALSE,]  
collapsed$dup <- NULL  
  
collapsed$antimicrobial <- NA  
collapsed$antimicrobial[grepl('Klebsiella pneumoniae', collapsed$org_name, ignore.case  
= T)] <- '3rd gen cephalosporins'  
collapsed$antimicrobial[grepl('Streptococcus pneumoniae', collapsed$org_name,  
ignore.case = T)] <- 'Penicillins'  
collapsed$antimicrobial[grepl('Staphylococcus aureus', collapsed$org_name, ignore.case  
= T)] <- 'Methicillin'  
collapsed$antimicrobial[grepl('Escherichia coli', collapsed$org_name, ignore.case = T)]  
<- '3rd gen cephalosporins'  
  
collapsed$source <- 'MLW'  
collapsed$country <- 'MWI'  
collapsed$location_id <- 43  
  
collapsed$age_group_id <- NA  
collapsed$age_group_name <- NA  
collapsed$age_group_id[collapsed$age<5] <- 1  
collapsed$age_group_name[collapsed$age<5] <- 'Under 5'  
collapsed$age_group_id[collapsed$age>=5 & collapsed$age <15] <- 23  
collapsed$age_group_name[collapsed$age>=5 & collapsed$age <15] <- '5-14 years'  
collapsed$age_group_id[collapsed$age>=15 & collapsed$age <49] <- 24  
collapsed$age_group_name[collapsed$age>=15 & collapsed$age <49] <- '15-49 years'  
collapsed$age_group_id[collapsed$age>=50 & collapsed$age <70] <- 25  
collapsed$age_group_name[collapsed$age>=50 & collapsed$age <70] <- '50-69 years'  
collapsed$age_group_id[collapsed$age>=70] <- 26  
collapsed$age_group_name[collapsed$age>=70] <- '70+ years'  
  
collapsed$sex[is.na(collapsed$sex)] <- 'Unknown'  
collapsed$sex[collapsed$sex == ""] <- 'Unknown'  
collapsed$sex_id <- NA  
collapsed$sex_id[grepl('Male', collapsed$sex, ignore.case = T)] <- 1
```

```
collapsed$sex_id[grepl('Female', collapsed$sex, ignore.case = T)] <- 2
collapsed$sex_id[grepl('unknown', collapsed$sex, ignore.case = T)] <- NA
```



```
clean_data <- collapsed[,.(location_id,
                             country,
                             year_id,
                             age,
                             age_group_id,
                             age_group_name,
                             sex_id,
                             sex,
                             specimen_id,
                             specimen_source,
                             pathogen = org_name,
                             antimicrobial,
                             resistance,
                             source)]
```

```
saveRDS(clean_data, 'FILEPATH')
```

```
year_site <- clean_data[,.(n_resistant = sum(resistance),
                           val = round(sum(resistance)/length(resistance),2),
                           sample_size = length(resistance),
                           variance = ((sum(resistance)/length(resistance))*(1-
(sum(resistance)/length(resistance))))/length(resistance),
                           sex_id = 3,
                           age_id = 22,
                           measure = 'proportion',
                           is_outlier = 0),
                           c("location_id", "country", "year_id", "pathogen",
'antimicrobial', 'source')]
```

```
year_site$year_id <- as.factor(year_site$year_id)
```

```
png('FILEPATH',
     height = 20, width = 25, units = 'cm', res = 150)
ggplot(year_site)+
  geom_boxplot(aes(x = year_id, y = val, colour = pathogen))+
  facet_wrap(~pathogen)+
  theme_bw()+
  labs(x = 'Year', y = 'Resistance (proportion)', colour = 'Pathogen') +
  theme(legend.position="bottom")
dev.off()
```

```
write.csv(year_site, 'FILEPATH', row.names = F)
```

```
all_aggregate <- clean_data[,.(n_resistant = sum(resistance),
                              val = round(sum(resistance)/length(resistance),2),
                              sample_size = length(resistance),
                              variance = ((sum(resistance)/length(resistance))*(1-
(sum(resistance)/length(resistance))))/length(resistance),
                              sex_id = 3,
                              age_id = 22,
```

```

        measure = 'proportion',
        is_outlier = 0),
    c("location_id", "country", "year_id", "pathogen",
'antimicrobial', 'source')])

```

```

png('FILEPATH',
    height = 10, width = 15, units = 'cm', res = 150)
ggplot(all_aggregate)+
  geom_bar(aes(x = year_id, y = val, fill = pathogen), stat = 'identity',
position=position_dodge())+
  theme_bw()+
  labs(x = 'Year', y = 'Resistance (%)', fill = 'Pathogen') +
  theme(legend.position="bottom")
dev.off()

```

```

write.csv(all_aggregate, 'FILEPATH', row.names = F)

```

```

by_source <- clean_data[,.(n_resistant = sum(resistance),
        sample_size = length(resistance),
        prop_resistant =
round(sum(resistance)/length(resistance),2)),
        c("year_id", "pathogen", "specimen_source")]

```

```

png('FILEPATH',
    height = 10, width = 15, units = 'cm', res = 150)
ggplot(by_source)+
  geom_bar(aes(x =year_id, y = prop_resistant, fill = specimen_source), stat =
'identity', position=position_dodge())+
  theme_bw()+
  scale_fill_brewer(palette = "Dark2")+
  labs(x = 'Year', y = 'Resistance (proportion)', fill = 'specimen_source')+
  facet_wrap(~pathogen) +
  theme(legend.position="bottom")
dev.off()

```

```

by_sex <- clean_data[,.(n_resistant = sum(resistance),
        sample_size = length(resistance),
        prop_resistant = round(sum(resistance)/length(resistance),2)),
        c("year_id", "pathogen", "sex")]

```

```

by_sex <- by_sex[!is.na(by_sex$sex),]

```

```

png('FILEPATH',
    height = 10, width = 15, units = 'cm', res = 150)
ggplot(by_sex)+
  geom_bar(aes(x =year_id, y = prop_resistant, fill = sex), stat = 'identity',
position=position_dodge())+
  theme_bw()+
  scale_fill_brewer(palette = "Dark2")+
  labs(x = 'Year', y = 'Resistance (%)', fill = 'Sex')+
  facet_wrap(~pathogen) +
  theme(legend.position="bottom")

```

```
dev.off()
```

```
by_age_group <- clean_data[,.(n_resistant = sum(resistance),  
                             sample_size = length(resistance),  
                             prop_resistant =  
round(sum(resistance)/length(resistance),2)),  
                           c("year_id", "pathogen", "age_group_name")]
```

```
by_age_group <- by_age_group[!is.na(by_age_group$age_group_name),]
```

```
png('FILEPATH',  
    height = 10, width = 15, units = 'cm', res = 150)  
ggplot(by_age_group)+  
  geom_bar(aes(x =year_id, y = prop_resistant, fill = age_group_name), stat =  
'identity', position=position_dodge())+  
  theme_bw()+  
  scale_fill_brewer(palette = "Dark2")+  
  labs(x = 'Year', y = 'Resistance (%)', fill = 'Age group')+  
  facet_wrap(~pathogen) +  
  theme(legend.position="bottom")  
dev.off()
```



```
by_source2 <- dcast(by_source, year_id + pathogen ~ specimen_source, value.var =  
'prop_resistant')  
by_age_group2 <- dcast(by_age_group, year_id + pathogen ~ age_group_name, value.var =  
'prop_resistant')  
by_sex2 <- dcast(by_sex, year_id + pathogen ~ sex, value.var = 'prop_resistant')
```

```
by_source2 <- by_source2[,1:3]  
by_source <- merge(by_source2, by_source, by = c('year_id', 'pathogen'), all.x = T,  
all.y = T)
```

```
by_age_group2 <- by_age_group2[,1:3]  
by_age_group <- merge(by_age_group2, by_age_group, by = c('year_id', 'pathogen'), all.x  
= T, all.y = T)
```

```
by_sex2 <- by_sex2[,1:3]  
by_sex <- merge(by_sex2, by_sex, by = c('year_id', 'pathogen'), all.x = T, all.y = T)
```

```
sex_stats <- by_sex[,.(mean_estimate = round(t.test(Female,  
prop_resistant)$estimate[2],2),  
                      p.value = round(t.test(Female, prop_resistant)$p.value,4),  
                      sample_size = sum(sample_size)),  
                    by = c('pathogen', 'sex'))]
```

```
colnames(sex_stats)[2] <- 'variable'
```

```
age_group_stats <- by_age_group[,.(mean_estimate = round(t.test(`15-49 years`,  
prop_resistant)$estimate[2],2),  
                                   p.value = round(t.test(`15-49 years`,  
prop_resistant)$p.value,4),  
                                   sample_size = sum(sample_size)),  
                                by = c('pathogen', 'age_group_name')]  
colnames(age_group_stats)[2] <- 'variable'
```



```

stats <- rbind(sex_stats, age_group_stats)
write.csv(stats, 'FILEPATH', row.names = F)

by_source <- clean_data[,.(n_resistant = sum(resistance),
                           sample_size = length(resistance),
                           prop_resistant =
round(sum(resistance)/length(resistance),2)),
                           c("year_id", "pathogen", "specimen_source")]

by_sex <- clean_data[,.(n_resistant = sum(resistance),
                       sample_size = length(resistance),
                       prop_resistant = round(sum(resistance)/length(resistance),2)),
                       c("year_id", "pathogen", "sex")]

by_sex <- by_sex[!is.na(by_sex$sex),]

by_age_group <- clean_data[,.(n_resistant = sum(resistance),
                              sample_size = length(resistance),
                              prop_resistant =
round(sum(resistance)/length(resistance),2)),
                              c("year_id", "pathogen", "age_group_name")]

by_age_group <- by_age_group[!is.na(by_age_group$age_group_name),]

by_source2 <- dcast(by_source, year_id + pathogen ~ specimen_source, value.var =
'prop_resistant')
by_age_group2 <- dcast(by_age_group, year_id + pathogen ~ age_group_name, value.var =
'prop_resistant')
by_sex2 <- dcast(by_sex, year_id + pathogen ~ sex, value.var = 'prop_resistant')

by_source2 <- by_source2[,1:4]
by_source <- merge(by_source2, by_source, by = c('year_id', 'pathogen'), all.x = T,
all.y = T)

by_age_group2 <- by_age_group2[,1:4]
by_age_group <- merge(by_age_group2, by_age_group, by = c('year_id', 'pathogen'), all.x
= T, all.y = T)

by_sex2 <- by_sex2[,1:4]
by_sex <- merge(by_sex2, by_sex, by = c('year_id', 'pathogen'), all.x = T, all.y = T)

sex_stats <- by_sex[,.(mean_estimate = round(t.test(Female,
prop_resistant)$estimate[2],2),
                      p.value = round(t.test(Female, prop_resistant)$p.value,4),
                      sample_size = sum(sample_size)),
                      by = c('pathogen', 'sex'))

colnames(sex_stats)[2] <- 'variable'

age_group_stats <- by_age_group[,.(mean_estimate = round(t.test(`15-49 years`,
prop_resistant)$estimate[2],2),
                                p.value = round(t.test(`15-49 years`,
prop_resistant)$p.value,4),
                                sample_size = sum(sample_size)),
                                by = c('pathogen', 'age_group_name'))

```

```
by = c('pathogen', 'age_group_name')]
```

```
colnames(age_group_stats)[2] <- 'variable'
```

```
stats <- rbind(sex_stats, age_group_stats)  
write.csv(stats, 'FILEPATH', row.names = F)
```