Applying the BEST-COST code to German END data

This is to gather experience using the *healthiar* package.

```
rm(list=ls())
library(healthiar)
library(tidyverse)
library(readxl)
library(janitor)
```

Reading data

To this end, I would like to apply the code to real data created in Hessen (a part of Germany) by the agency HLNUG for the official END noise mapping and compare to other calculations.

Reading external data

The population of Hessen as published here: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugJ/31122021_Auszug_GV.xlsx?__blob=publicationFile&v=4

```
population_Hessen <- read_excel("data/Population Hessen.xlsx") %>%
  filter(!is.na(Gem)) %>%
  select(!Fläche) %>%
  mutate(gemeindekennziffer=paste0(RB,Kreis,Gem) %>%
           as.numeric,.keep="unused",.before=1) %>%
#  mutate(Gemeindename=str_remove(Gemeindename,",.*"))
    mutate(Gemeindename=NULL,population=Bevölkerung,.keep="unused")
```

Reading END results

Reading END result data from https://www.hlnug.de/fileadmin/dokumente/laerm/laermkartierung/ULK2022StatistikGesamt_20240704.xlsx. The data can be shown like this:

```
agglomerations<-c("Darmstadt","Frankfurt am Main","Offenbach am Main","Wiesbaden","Hanau","Kassel")

road <- read_excel("data/ULK2022StatistikGesamt_20240704.xlsx",
    sheet = "Straßenlärm", skip = 1) %>%
  janitor::clean_names() %>%
  mutate(gemeinde_kennziffer=as.numeric(gemeinde_kennziffer))%>%
  filter(!is.na(gemeinde_kennziffer)) %>% #remove sum rows
  select(gemeinde_kennziffer:anzahl_belasteter_l_night_ab_70|starts_with("zahl")) %>%
```

```
  rename_with(~str_remove(.x, "anzahl_belasteter_"), everything()) %>%
  rename_with(~str_remove_all(.x, "_"), everything()) %>%
  rename_with(~str_replace(.x, "ab","|"), everything()) %>%
  left_join(population_Hessen)


road_long<-road%>%
  pivot_longer(
    cols = starts_with("l"),
    values_to = "exposed"
  ) %>%
  separate_wider_delim(
    col = name,
    delim = "|",
    names = c("measure", "range")
  )%>%
  separate_wider_delim(
    col = range,
    delim = "bis",
    names = c("fromdB", "todB"),
    too_few = "align_start"
  )%>%
  mutate(fromdB=as.numeric(fromdB),todB=as.numeric(todB)) %>%
  mutate(
    centraldB=fromdB+2,
    .after=4
  ) %>%
  mutate(agglomeration = (namestadtgemeinde %in% agglomerations))

hessen_exp_road<-road_long %>%
  group_by(measure,fromdB,centraldB,todB,agglomeration) %>%
  summarise(exposed = sum(exposed, na.rm = TRUE))%>%
  pivot_wider(
    names_from = measure,
    values_from = exposed
  )

hessen_exp_road
```

Calculate Impact

Annoyance and sleep disturbance

To calculate HA and HSD with *healhtiar* is straight forward.

```
ha <- function(level_exposure,number_exposed) {
  result<-attribute_health(
    approach_risk = "absolute_risk",
    exp_central = level_exposure,
    pop_exp = number_exposed,
    erf_eq_central = "78.9270-3.1162*c+0.0342*c^2"
  )
  return(result$health_main$impact)
}

# hessen_exp_road %>%
#   filter(agglomeration==FALSE) %>%
#   {ha(.$centraldB,.$lden)}


haList <- function(disag_ID,ag_ID,level_exposure,number_exposed,pop) {
  result<-attribute_health(
    geo_id_disaggregated = disag_ID,
    geo_id_aggregated = ag_ID,
```

```
    approach_risk = "absolute_risk",
    exp_central = as.list(level_exposure),
    pop_exp = as.list(number_exposed),
    population = as.list(pop),
    erf_eq_central = "78.9270-3.1162*c+0.0342*c^2"
  )
  return(result$health_detailed)
}


resDEN<-road_long %>%
  filter(measure=="lden") %>%
#  filter(agglomeration==FALSE) %>%
  {haList(.$namestadtgemeinde,.$agglomeration,.$centraldB,.$exposed,.$population)} %>%
  .$impact_raw %>%
  mutate(outcome="HA")

hsdList <- function(disag_ID,ag_ID,level_exposure,number_exposed,pop) {
  result<-attribute_health(
    geo_id_disaggregated = disag_ID,
    geo_id_aggregated = ag_ID,
    approach_risk = "absolute_risk",
    exp_central = as.list(level_exposure),
    pop_exp = as.list(number_exposed),
    population = as.list(pop),
    erf_eq_central = "19.4312 - 0.9336*c + 0.0126 * c^2"
  )
  return(result$health_detailed)
}

resNight<-road_long %>%
  filter(measure=="lnight") %>%
#  filter(agglomeration==FALSE) %>%
  {hsdList(.$namestadtgemeinde,.$agglomeration,.$centraldB,.$exposed,.$population)} %>%
  .$impact_raw %>%
  mutate(outcome="HSD")
res<-bind_rows(resDEN,resNight) %>%
  select(starts_with("geo"), exp, impact, outcome) %>%
  mutate(geo_id_aggregated=as.logical(geo_id_aggregated)) %>%
  group_by(outcome,geo_id_disaggregated) %>%
  summarise(impact=sum(impact)) %>%
  pivot_wider(names_from=outcome,values_from = impact)

## `summarise()` has grouped output by 'outcome'. You can override using the
## `.groups` argument.

res
```

Heart diseases

In order to calculate attributable IHD-numbers, I still need the used prevalence data used in the HLNUG calculations.

Compare

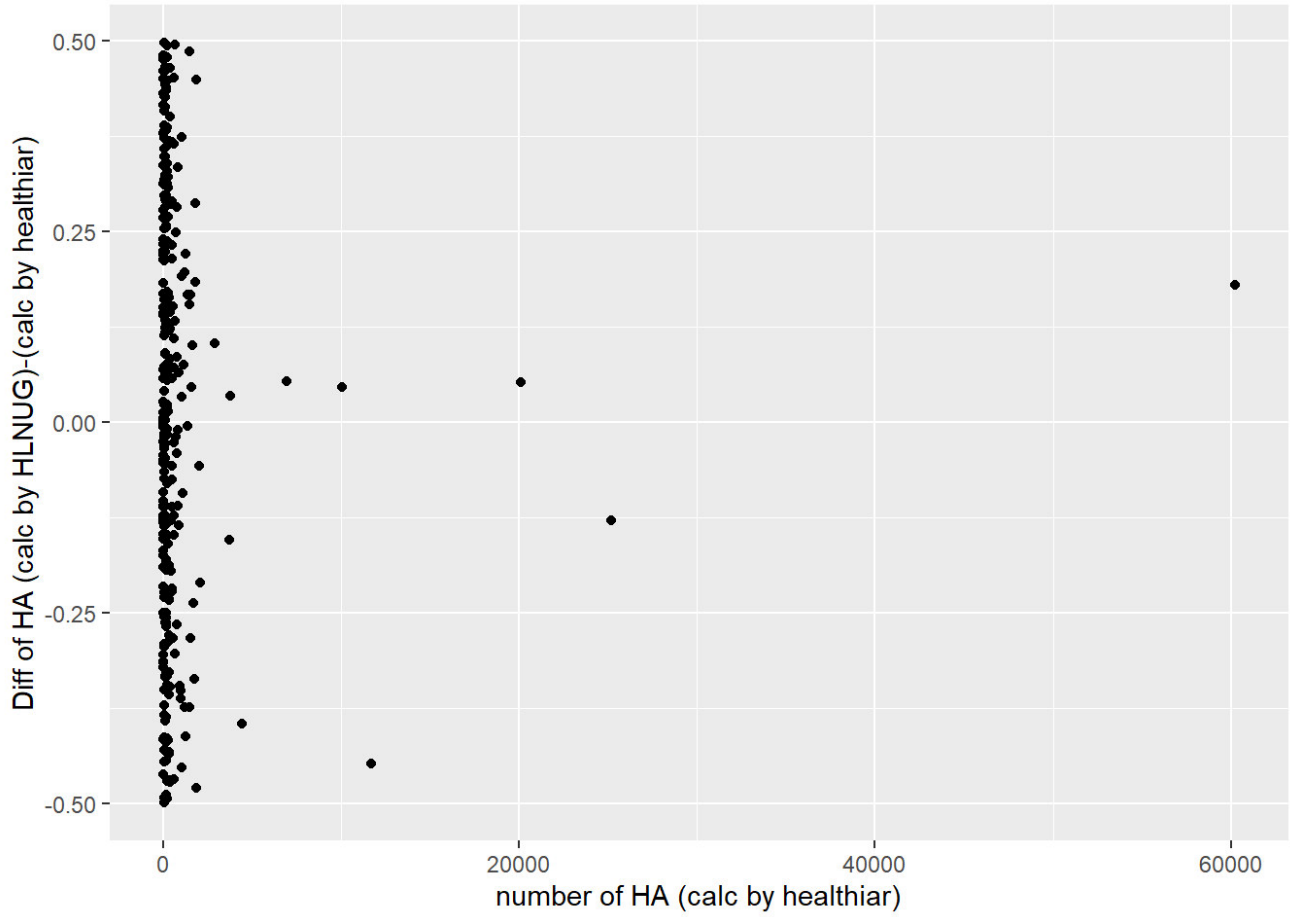
Here, I perform a validity-check, by comparing the calculated impact of each *Gemeinde* (municipality) calculated by *healthiar*. vs. HLNUG results.

Number of highly annoyed

The differences between the HA-numbers calculated account to always less than $\pm 0,5$. This seems to be only due to rounding differences.

```
comp<-res %>%
  left_join(road,by=c(
    "geo_id_disaggregated"="namestadtgemeinde"
  )) %>%
  mutate(diffHA=zahlderfallestarkerbelastigung - HA) %>%
  mutate(diffHSD=zahlderfallestarkerschlafstorung - HSD)

comp %>%
  ggplot(aes(x=HA, y=diffHA))+
  geom_point() +
  xlab("number of HA (calc by healthiar)") +
  ylab("Diff of HA (calc by HLNUG)-(calc by healthiar)")
```



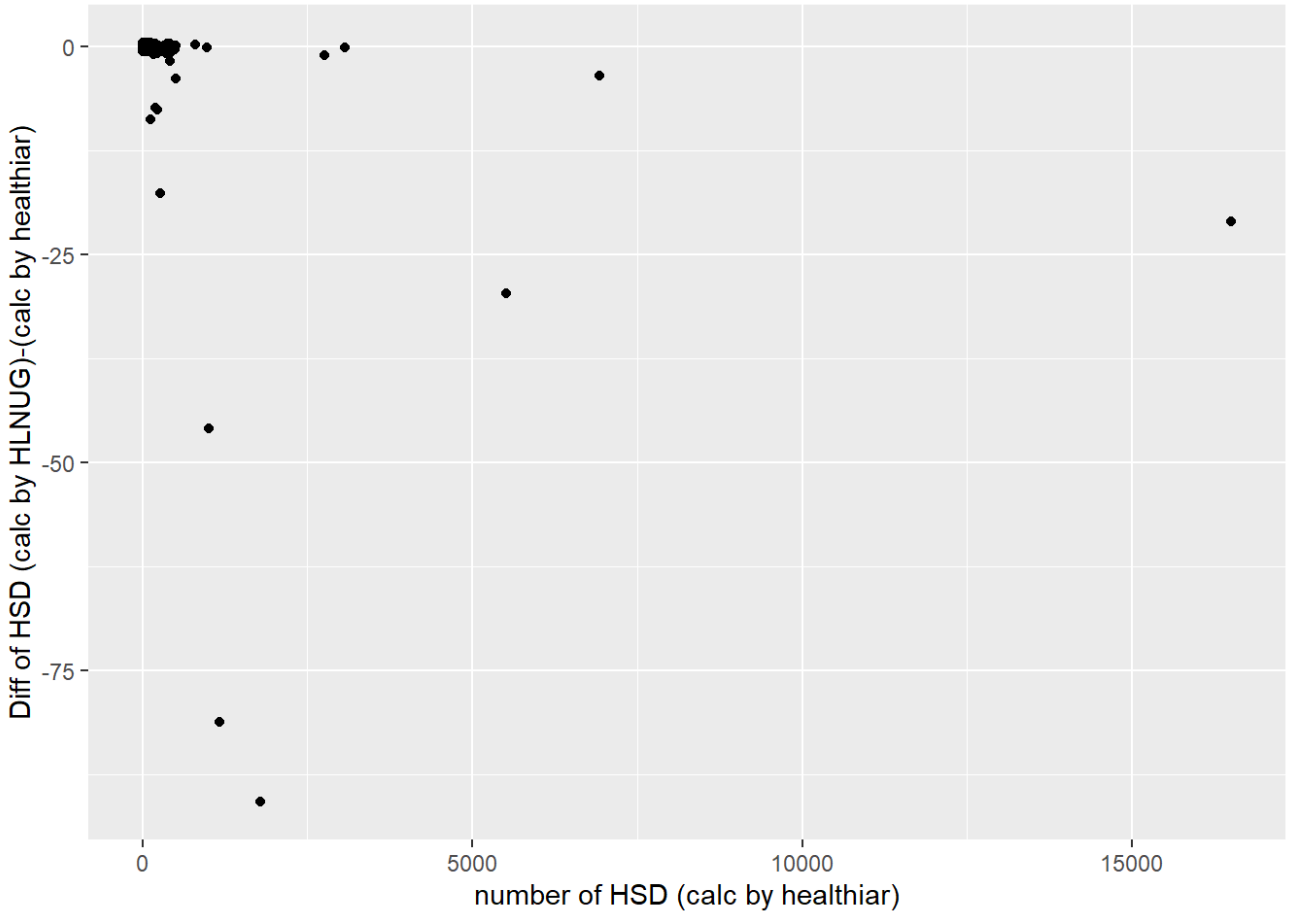
Each point in the plot accounts for one municipality. The total difference in HA amounts to 2.857592, which is totally acceptable.

Number of highly sleep disturbed

```
casesOKHSD<-comp %>% filter(abs(diffHSD)< 0.5) %>% count() %>% unlist
casesbigDiffHSD<-comp %>% filter(diffHSD< -0.5) %>% count() %>% unlist
```

The picture of HSD-numbers is more diverse: The differences between the HSD-numbers calculated account to always less than $\pm 0,5$ in most (407) cases, however in 19 cases, the difference between the two calculations amounts to up to a difference of 90 for a single municipality.

```
comp %>%
  ggplot(aes(x = HSD, y = diffHSD)) + geom_point() +
  xlab("number of HSD (calc by healthiar)") +
  ylab("Diff of HSD (calc by HLNUG)-(calc by healthiar)")
```



The total difference in HSD amounts to -325.078388, which seems quote a lot. However it amounts to a fraction of -0.0050372 of all HSD, which is still acceptable.