Crop Analysis Q3 2013

nbngo

2021-12-02

```
# notes from last run----------------------



# 1. setup ---------




## 1.1. load functions -----------------

# install.packages("here")
library(here)
```

```
## here() starts at C:/Users/nbngo/OneDrive/PhD/wal-water
```

```
source(here("3 Scripts", "general_functions.R"))

loadpackage('dplyr')
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
loadpackage('ggplot2')
```

```
## Loading required package: ggplot2
```

```
loadpackage('reshape2')
```

```
## Loading required package: reshape2
```

```
loadpackage('raster')
```

```
## Loading required package: raster

## Loading required package: sp

##
## Attaching package: 'raster'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
loadpackage('sf')
```

```
## Loading required package: sf

## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1
```

```
## 1.2 load data --------

### data folder
rdir <- '2 Data/1 Raw'
pdir <- '2 Data/2 Processed'

### survey data

surv14 <- read.csv(file = here(pdir, "UtiSurv_2014_AWalCEHD_Wal/Survey2014_obs_AquaWal_prd.csv"))

### price data

price <- read.csv(file = here(pdir, 'Water_price_AWal_Wal/water_price_Wal_12_17.csv'))


### location
load(file= here(pdir, 'UtiSurv_2014_AWalCEHD_Wal/Addresses/surv14_coord_PICC.Rdata'))

### urbanization

urban_10 <- raster(here(pdir, "Urban_5cat_Ahmed_Wal/LU2010_5cls_x25.flt"))

crs(urban_10) <- st_crs(surv14_coords)

# 2. merge data ----------

surv14$year <- 2014

df <- dplyr::left_join(surv14, price)[, c('id', 'CVD', 'CVA', 'csmptv', 'inceqa', 'rwtank', 'hhs_0_19',
```

```
## Joining, by = c("dtbtor", "year")
```

```
df$hhs <- df$hhs_0_19 + df$hhs_20_95

df$csmptv[df$csmptv > 300] <- NA



urban <- cbind.data.frame(id = surv14_coords$id, urban = extract(urban_10, surv14_coords))

urban <- urban[urban$urban > -1,]

df <- inner_join(df, urban)
```

```
## Joining, by = "id"
```

```
df$cspc <- df$csmptv/df$hhs

df <- df[!is.na(df$csmptv) & !is.na(df$inceqa) & !is.na(df$CVD),]

df <- df[df$dtbtor %in% c("CILE", "SWDE", "IECBW"),]

df$ab30 <- as.numeric(df$csmptv > 30)



# 3. water bills ----
## 3.1. current -----
```

```
df$bill_cur <- (20*df$CVD + 30*df$CVA) + 0.0125*df$csmptv + 0.5*df$csmptv*df$CVD + 0.5*(df$csmptv - 30)


unique(df[, c("dtbtor", "CVA")])
```

```
##    dtbtor   CVA
## 1    SWDE 1.745
## 2    CILE 1.745
## 11  IECBW 1.745
```

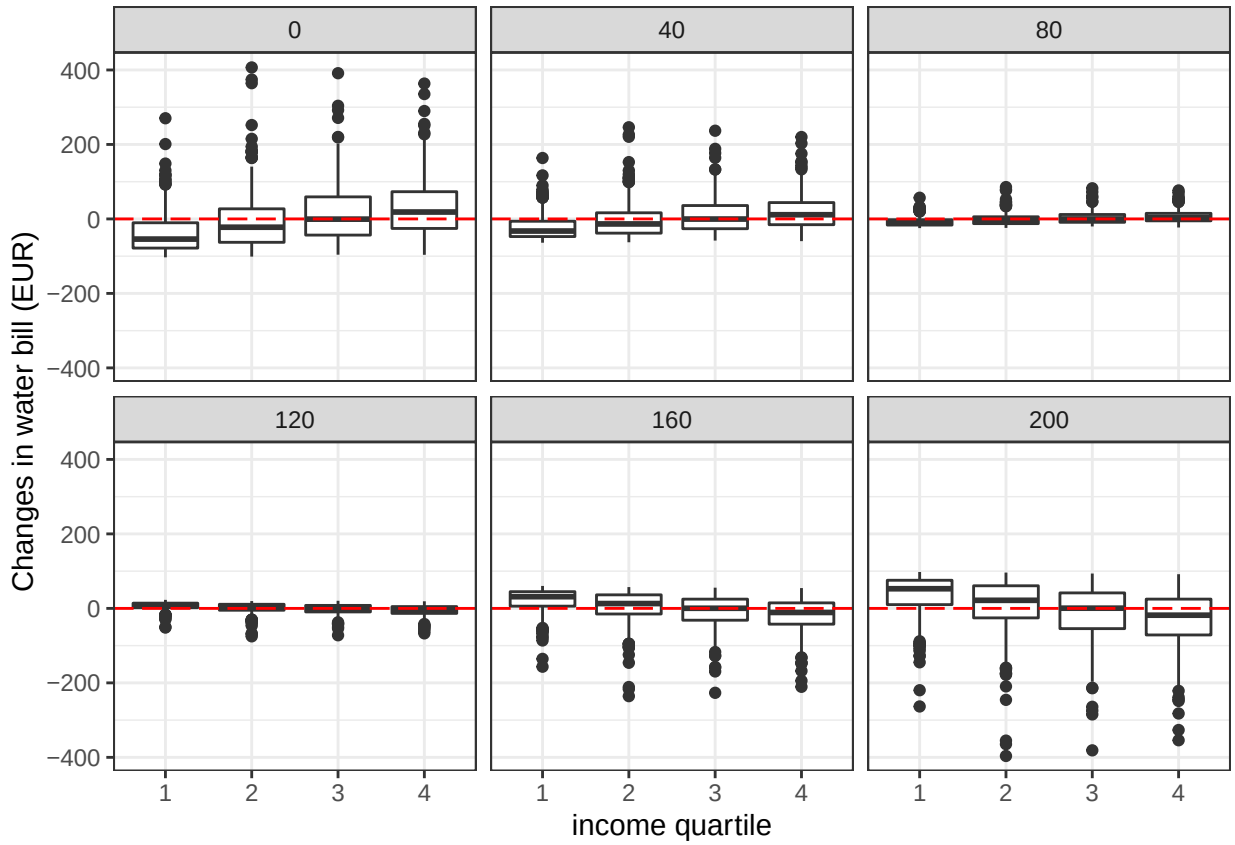
```
## 3.2. current with different fixed  -----
```

```
vary_fixed_f <- function(dtbtorname = "SWDE", fixeds = seq (0, 200, 40)) {
  tmpdf <- df[df$dtbtor %in% dtbtorname,]

  total <- sum(tmpdf$bill_cur)
  fse <- sum(0.0125*tmpdf$csmptv)
  cvd_v <- numeric()

  for (fixed in fixeds) {

    cvd <- (total - fse - fixed*nrow(tmpdf) - sum((tmpdf$csmptv - 30)*tmpdf$CVA*tmpdf$ab30))/sum(0.5*tm
    cvd_v <- c(cvd_v, cvd)
```

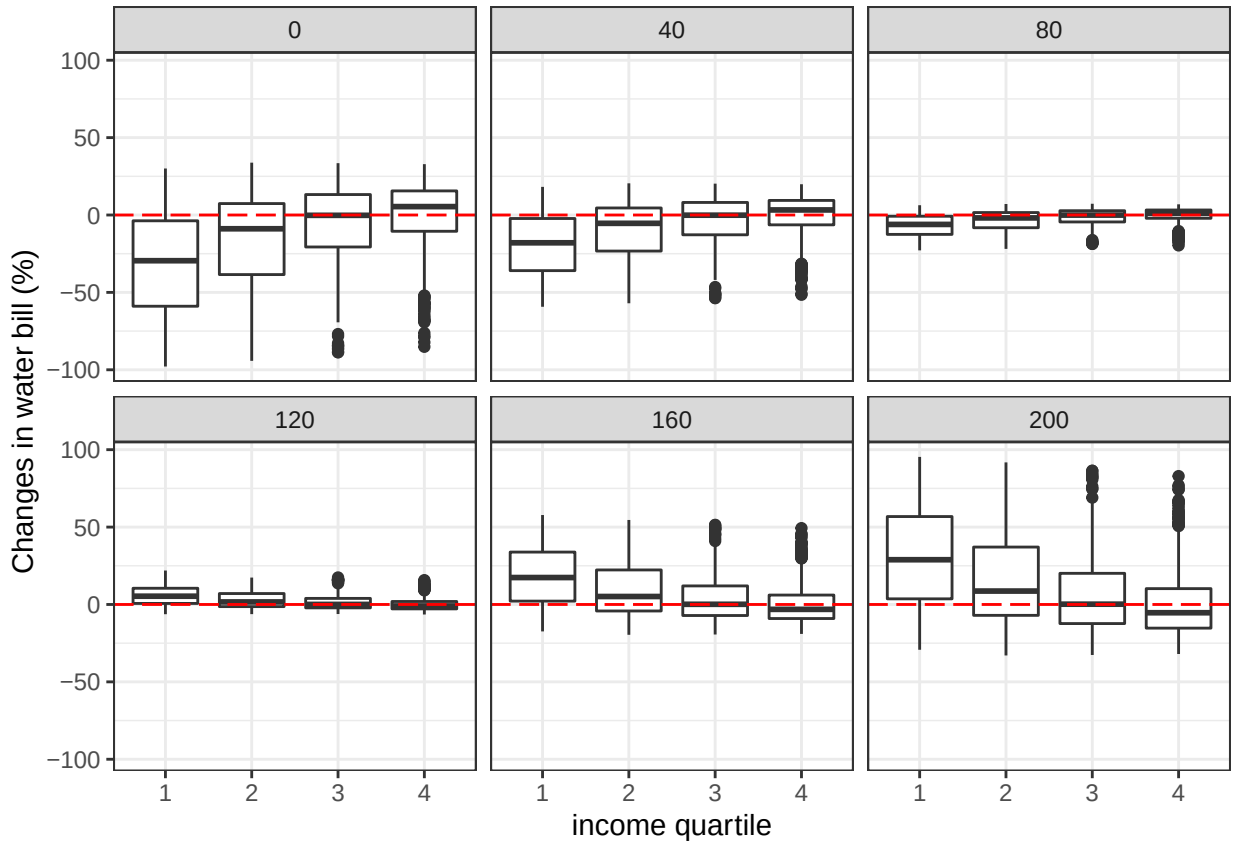
```
    tmpdf$bill_new <- fixed + 0.0125*tmpdf$csmptv + 0.5*tmpdf$csmptv*cvd + 0.5*(tmpdf$csmptv - 30)*cvd*
    tmpdf$diff <- tmpdf$bill_new - tmpdf$bill_cur
    tmpdf$difpct <- tmpdf$diff*100/tmpdf$bill_cur
    colnames(tmpdf)[ncol(tmpdf) - 2] <- paste("bill_fixed", fixed, sep = "_")
    colnames(tmpdf)[ncol(tmpdf) - 1] <- paste("difab_fixed", fixed, sep = "_")
    colnames(tmpdf)[ncol(tmpdf)] <- paste("difpc_fixed", fixed, sep = "_")

  }
  list(tmpdf, cvd_v)
}


vary_fixed_ls <- lapply(c("SWDE", "CILE", "IECBW"),vary_fixed_f)

cvd <- sapply(vary_fixed_ls, function(x) x[[2]])

write.table(cvd, "clipboard", sep = "\t")

vary_fixed_df <- Reduce(rbind, lapply(vary_fixed_ls, function(x) x[[1]]))

df <- left_join(df, vary_fixed_df)
```

```
## Joining, by = c("id", "CVD", "CVA", "csmptv", "inceqa", "rwtank", "hhs_0_19", "hhs_20_95", "income",
```

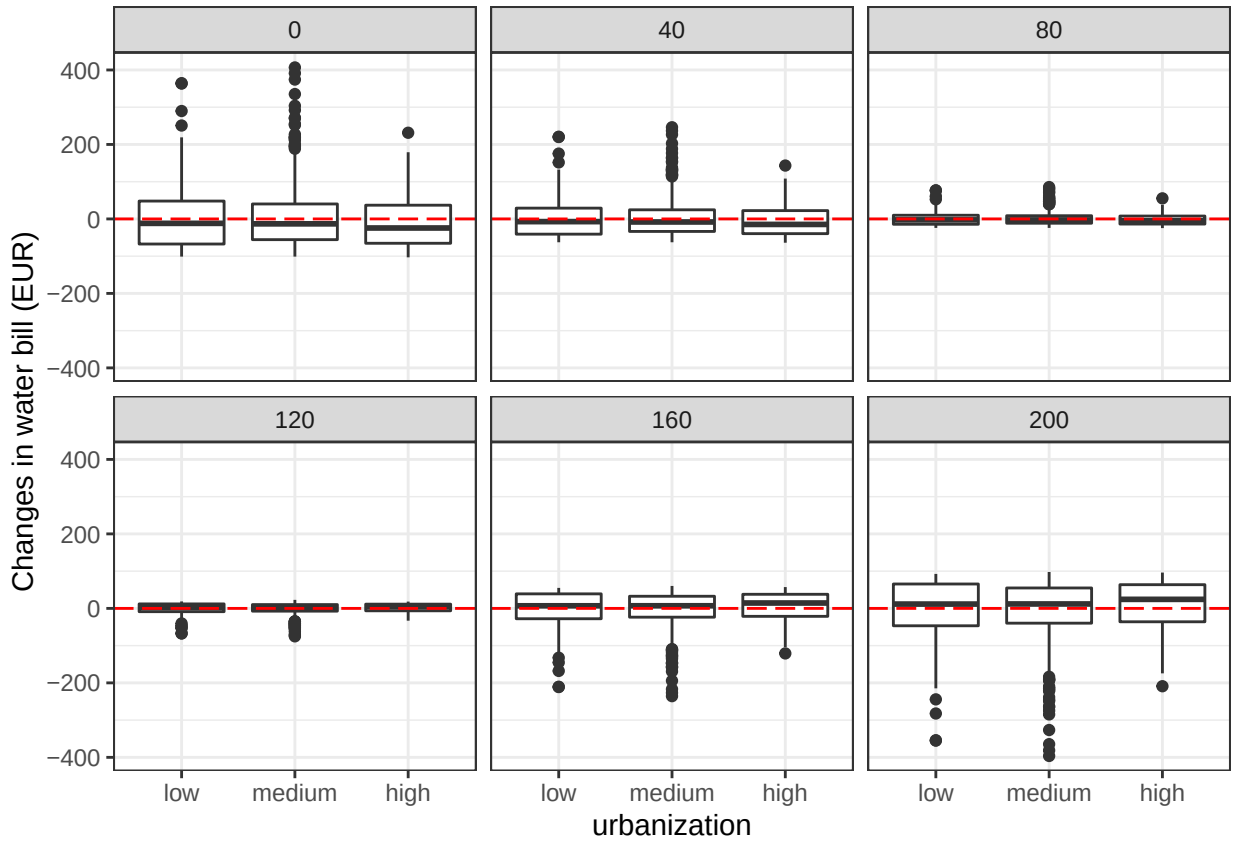
```
test <- df[,c("id", "income" , grep("difab_fixed", colnames(df), value = T))]

test <- test %>% mutate(incqnt = ntile(income, 4))
test$incqnt <- as.factor(test$incqnt)

test_lg <- melt(test, id.vars = c("id", "income", "incqnt"))
test_lg$fixed <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = incqnt, y = value)) +
  facet_wrap(.~ fixed) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "income quartile", y = "Changes in water bill (EUR)")
```



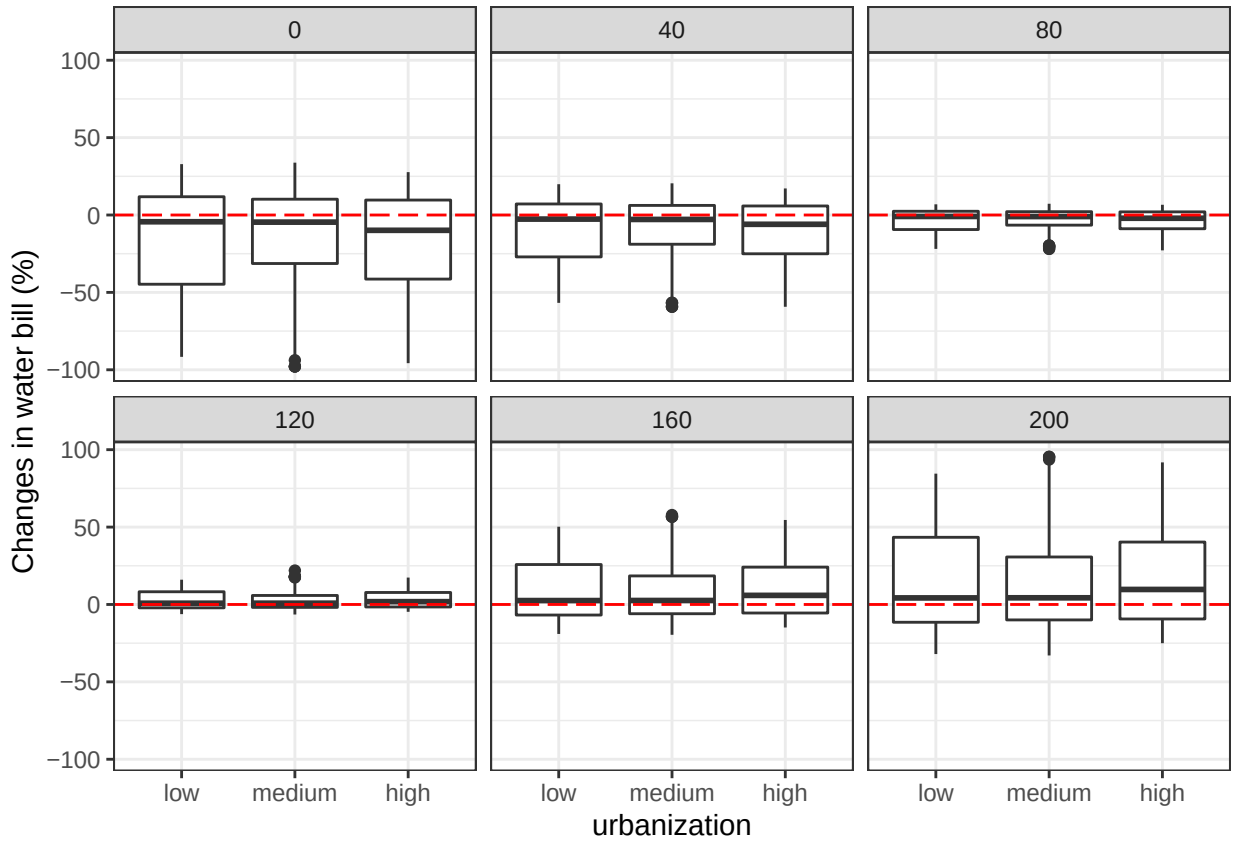

```
test <- df[,c("id", "income" , grep("difpc_fixed", colnames(df), value = T))]

test <- test %>% mutate(incqnt = ntile(income, 4))
test$incqnt <- as.factor(test$incqnt)

test_lg <- melt(test, id.vars = c("id", "income", "incqnt"))
test_lg$fixed <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = incqnt, y = value)) +
  facet_wrap(.~ fixed) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "income quartile", y = "Changes in water bill (%)")
```




```
df$urban <- factor(df$urban)
df$urban <- car::recode(df$urban, "c('0', '1') = 'low'; c('2','3') = 'medium'; c('4','5') = 'high' ")

df$urban <- factor(df$urban, levels = lvls(df$urban)[c(2,3,1)])

test <- df[,c("id", "urban" , grep("difab_fixed", colnames(df), value = T))]

test_lg <- melt(test, id.vars = c("id", "urban"))
test_lg$fixed <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = urban, y = value)) +
  facet_wrap(.~ fixed) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "urbanization", y = "Changes in water bill (EUR)")
```




```
test <- df[,c("id", "urban" , grep("difpc_fixed", colnames(df), value = T))]

test_lg <- melt(test, id.vars = c("id", "urban"))
test_lg$fixed <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = urban, y = value)) +
  facet_wrap(.~ fixed) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "urbanization", y = "Changes in water bill (%)")
```




```
# ## 3.3. linear with and without fixed -----
#
# unpr <- sum(df$bill_14)/sum(df$csmptu)
#
# df$bill_ln <- df$csmptu*unpr
#
# df$dif_ln <- df$bill_ln - df$bill_14
#
# #### with fixed
# fixed <- 1.06*(20*mean(df$CVD) + 30*mean(df$CVA))
#
# unpr1 <- (sum(df$bill_14) - nrow(df)*fixed)/sum(df$csmptu)
#
# df$bill_lnwf <- fixed + df$csmptu*unpr1
#
#
# df$dif_lnwf <- df$bill_lnwf - df$bill_14
#
#
#
## 3.4. block tariffs per captia -----

### Brussels scheme ------------

df$bl2 <- ifelse(df$cspc <= 30 & df$cspc > 15, 1, 0)
df$bl3 <- ifelse(df$cspc <= 60 & df$cspc >30, 1, 0)
```

```r
df$bl4 <- ifelse(df$cspc > 60, 1, 0)

summary(df$bl3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.3846  1.0000  1.0000
```

```r
tot <- sum(df$bill_cur)


bl_2_1 <- 3.7696/2.115  # 1.782317
bl_3_1 <- 5.5726/2.115 # 2.634799
bl_4_1 <- 8.1338/2.115 # 3.845768

bl1 <- (tot - 25.23*nrow(df))/sum((df$cspc + (df$cspc-15)*(bl_2_1 - 1)*df$bl2 + (df$cspc-30)*(bl_3_1 - l

df$bill_brx <- 25.23 + bl1*(df$cspc + (df$cspc-15)*(bl_2_1 - 1)*df$bl2 + (df$cspc-30)*(bl_3_1 - bl_2_1)*

summary(df$bill_brx)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   28.57  167.70  273.18  305.18  405.28 1630.66
```

```r
summary(df$bill_cur)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   102.5   169.1   272.4   305.2   396.0  1202.4
```

```r
sum(df$bill_brx)
```

```
## [1] 468142.9
```

```r
sum(df$bill_cur)
```

```
## [1] 468142.9
```

```r
#
# df$bill_brx <- ifelse(df$cspc <= 60 & df$cspc >30, 25.23 + (15*bl1 + 15*bl2 + (df$cspc - 30)*bl3)*df$
#
# df$bill_brx <- ifelse(df$cspc <= 30 & df$cspc > 15, 25.23 + (15*bl1 + (df$cspc - 15)*bl2)*df$hhs, df$
#
# df$bill_brx <- ifelse(df$cspc <=  15, 25.23 + (df$cspc*bl1)*df$hhs, df$bill_brx)
#
# sum(df$bill_brx)
# sum(df$bill_14)
#
# df$dif_brx <- df$bill_brx - df$bill_14
#
## 3.5. rainwater tank tax ---------
```

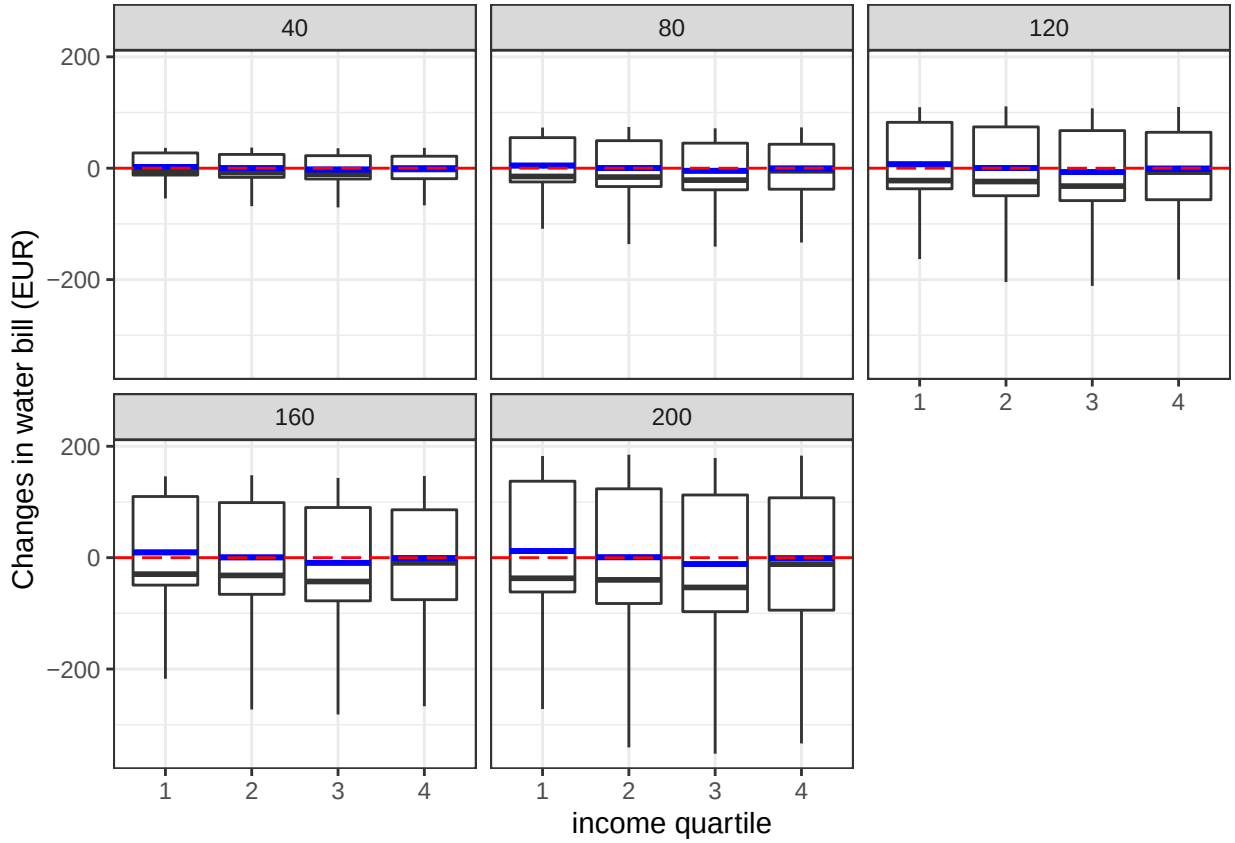
```r
df$rwtank <- as.numeric(df$rwtank %in% "yes")


vary_rwtf_f <- function(dtbtorname = "SWDE", rwtfs = seq (0, 200, 40)) {
  tmpdf <- df[df$dtbtor %in% dtbtorname,]

  total <- sum(tmpdf$bill_cur)
  fse <- sum(0.0125*tmpdf$csmptv)
  cvd_v <- numeric()

  for (rwtf in rwtfs) {

    rwtt <- sum(rwtf*tmpdf$rwtank)

    cvd <- (total - fse - rwtt - sum(30*tmpdf$CVA + (tmpdf$csmptv - 30)*tmpdf$CVA*tmpdf$ab30))/sum(20 +
    cvd_v <- c(cvd_v, cvd)
    tmpdf$bill_new <- rwtf*tmpdf$rwtank + 30*tmpdf$CVA + 20*cvd + 0.0125*tmpdf$csmptv + 0.5*tmpdf$csmpt
    tmpdf$diff <- tmpdf$bill_new - tmpdf$bill_cur
    tmpdf$difpct <- tmpdf$diff*100/tmpdf$bill_cur
    colnames(tmpdf)[ncol(tmpdf) - 2] <- paste("bill_rwtt", rwtf, sep = "_")
    colnames(tmpdf)[ncol(tmpdf) - 1] <- paste("difab_rwtt", rwtf, sep = "_")
    colnames(tmpdf)[ncol(tmpdf)] <- paste("difpc_rwtt", rwtf, sep = "_")

  }
  list(tmpdf, cvd_v)
}

### 0-200,40 -------------
vary_rwtf_ls <- lapply(c("SWDE", "CILE", "IECBW"),vary_rwtf_f)

cvd <- sapply(vary_rwtf_ls, function(x) x[[2]])

write.table(cvd, "clipboard", sep = "\t")

ndtb <- matrix(table(df$dtbtor)[c(3,1,2)])

write.table((cvd%*%ndtb*20 + unique(df$CVA)*30*1534)/1534, "clipboard", sep = "\t")

vary_rwtf_df <- Reduce(rbind, lapply(vary_rwtf_ls, function(x) x[[1]]))

df <- left_join(df, vary_rwtf_df)
```

```
## Joining, by = c("id", "CVD", "CVA", "csmptv", "inceqa", "rwtank", "hhs_0_19", "hhs_20_95", "income",
```

```r
test <- df[,c("id", "income" , grep("difab_rwtt", colnames(df), value = T))]

test <- test %>% mutate(incqnt = ntile(income, 4))
test$incqnt <- as.factor(test$incqnt)

test_lg <- melt(test, id.vars = c("id", "income", "incqnt"))
test_lg$rwtt <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg[test_lg$rwtt > 0,], aes(x = incqnt, y = value)) +
```

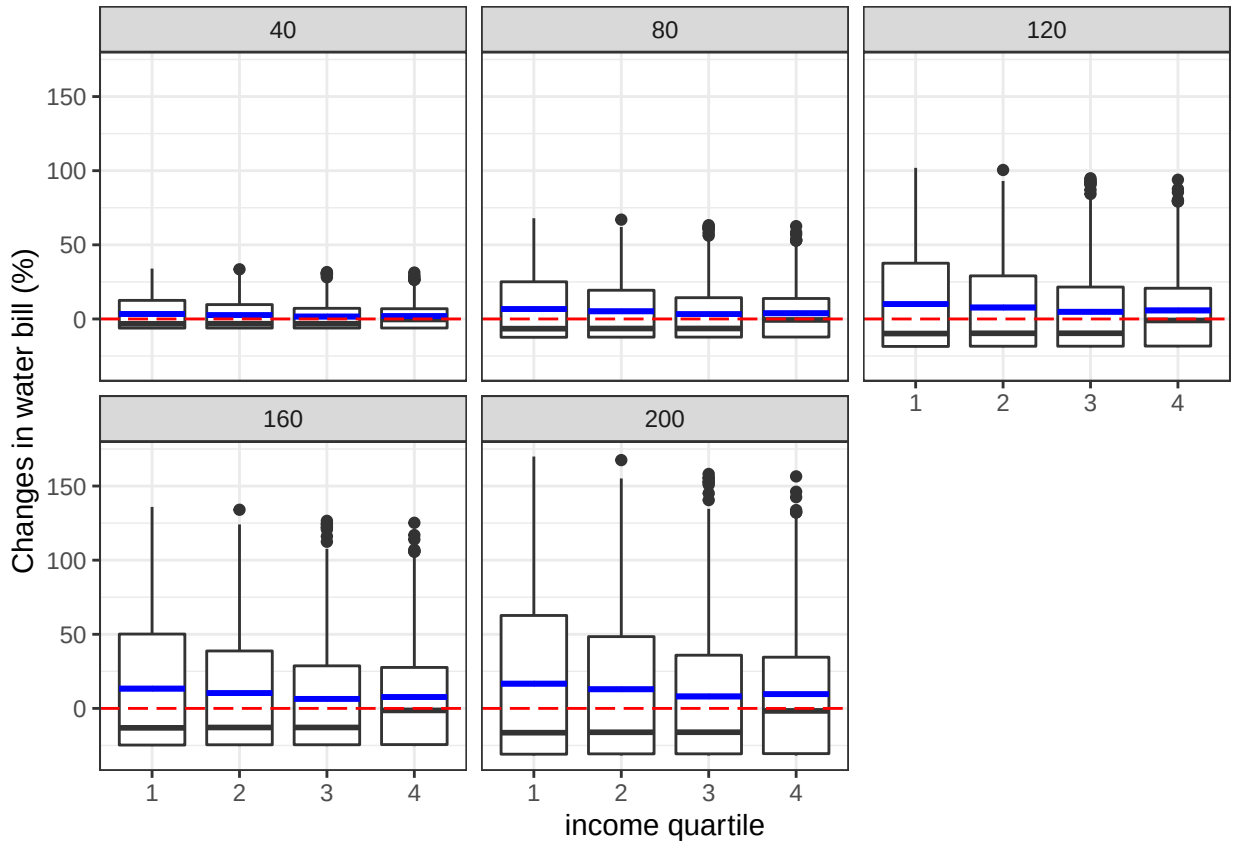
```
  geom_boxplot() +
  stat_summary(fun.y = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), col = 'blue', width = 0
  facet_wrap(.~ rwtt) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "income quartile", y = "Changes in water bill (EUR)")
```

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```

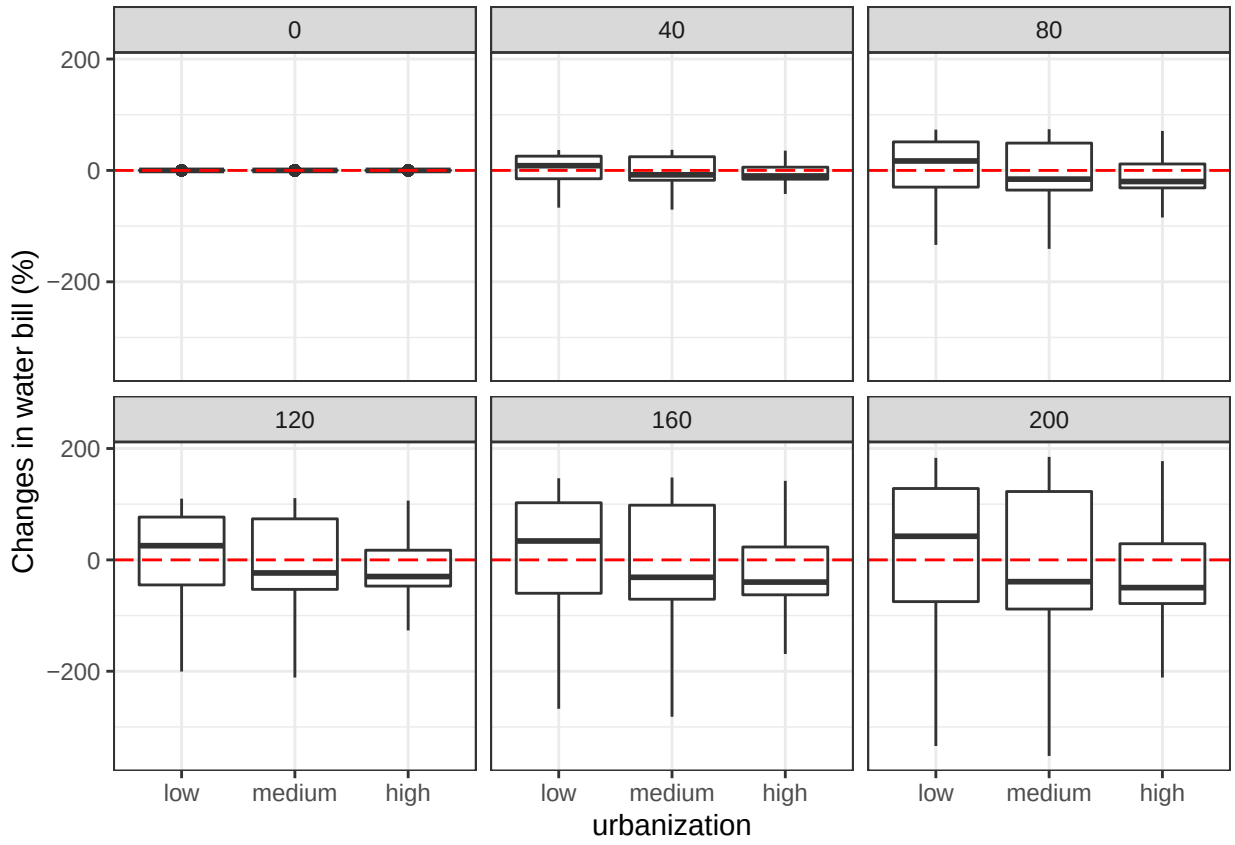



```
test <- df[,c("id", "income" , grep("difpc_rwtt", colnames(df), value = T))]

test <- test %>% mutate(incqnt = ntile(income, 4))
test$incqnt <- as.factor(test$incqnt)

test_lg <- melt(test, id.vars = c("id", "income", "incqnt"))
test_lg$rwtt <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg[test_lg$rwtt > 0,], aes(x = incqnt, y = value)) +
  geom_boxplot() +
  stat_summary(fun.y = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), col = 'blue', width = 0
  facet_wrap(.~ rwtt) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "income quartile", y = "Changes in water bill (%)")
```

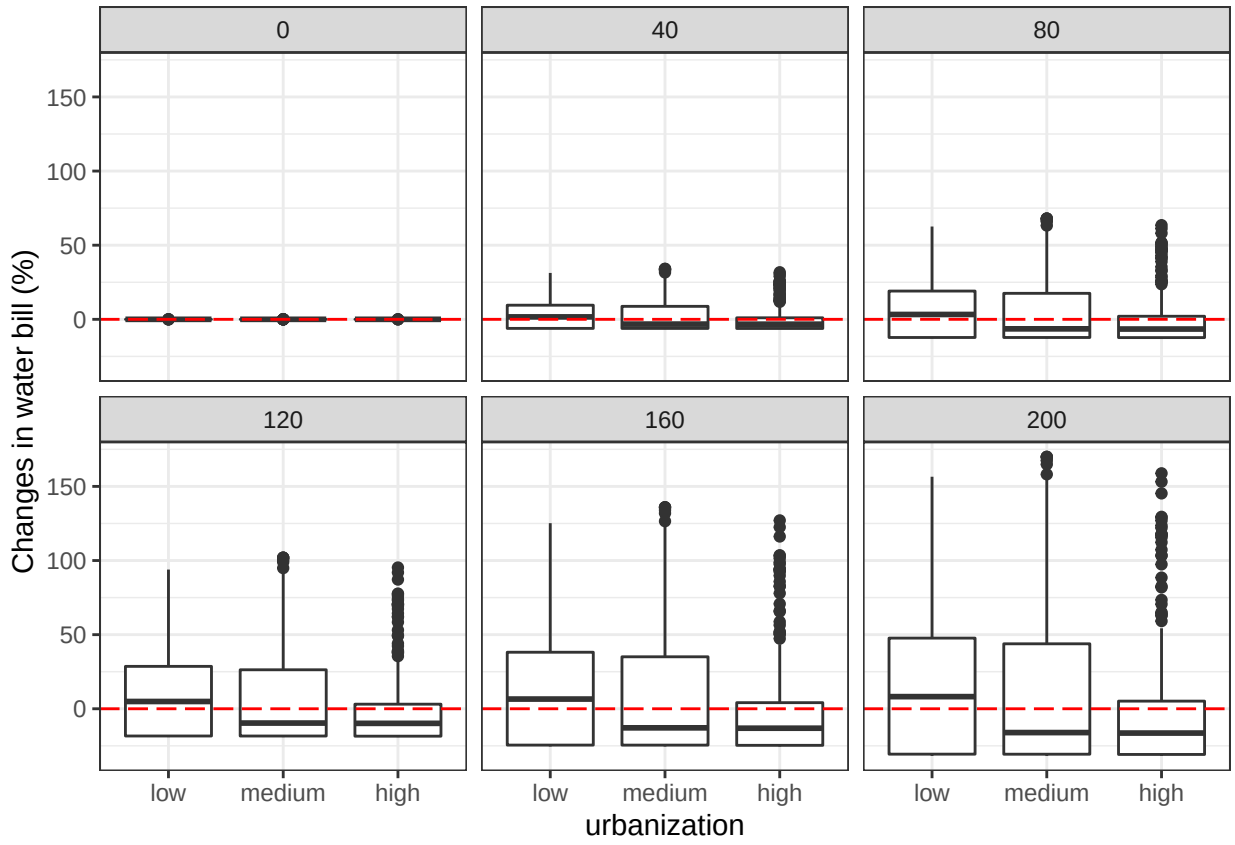
```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```




```
test <- df[,c("id", "urban" , grep("difab_rwtt", colnames(df), value = T))]

test_lg <- melt(test, id.vars = c("id", "urban"))
test_lg$rwtt <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = urban, y = value)) +
  facet_wrap(.~ rwtt) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "urbanization", y = "Changes in water bill (%)")
```




```
test <- df[,c("id", "urban" , grep("difpc_rwtt", colnames(df), value = T))]

test_lg <- melt(test, id.vars = c("id", "urban"))
test_lg$rwtt <- as.numeric(gsub(".*_", "", test_lg$variable))

ggplot(test_lg) +
  geom_boxplot(aes(x = urban, y = value)) +
  facet_wrap(.~ rwtt) +
  geom_hline(yintercept = 0, col = 'red', linetype = "longdash") +
  theme_bw() +
  labs(x = "urbanization", y = "Changes in water bill (%)")
```




```
# # 4. plots -----------
#
#
#
# ## 4.1. by income per capita -----
#
#
# df <- df %>% mutate(ipcqnt = ntile(inceqa, 10))
# df$ipcqnt <- as.factor(df$ipcqnt)
#
#
#
# df_sum <- df %>%
#   group_by(ipcqnt) %>%
#   summarise(count = n(),
#             avrbill = mean(bill_14, na.rm = T),
#             sdbill = sd(bill_14, na.rm = T),
#             rwtank = sum(rwtank %in% 'yes', na.rm = T)*100/n(),
#             avrhhs = mean(hhs, na.rm = T),
#             sdhhs = sd(hhs, na.rm = T),
#             avrinc = mean(inceqa, na.rm = T),
#             sdinc = sd(income, na.rm = T),
#             avrurban = mean(urban, na.rm = T),
#             sdurban = sd(urban, na.rm = T),
#             avr_ln = mean(bill_ln),
#             dif_ln = mean(dif_ln),
```

```
#             avr_lnwf = mean(bill_lnwf),
#             dif_lnwf = mean(dif_lnwf),
#             avr_brx = mean(bill_brx),
#             dif_brx = mean(dif_brx))
#
#
#
# df_sum$dif_ln_prc <- df_sum$dif_ln*100/df_sum$avrbill
#
# df_sum$dif_lnwf_prc <- df_sum$dif_lnwf*100/df_sum$avrbill
#
# df_sum$dif_brx_prc <- df_sum$dif_brx*100/df_sum$avrbill
#
#
# plotdf <- melt(df_sum[, c('ipcqnt', "dif_ln_prc", "dif_lnwf_prc", "dif_brx_prc")])
#
#
#
# ggplot(plotdf) +
#   geom_col(aes(x = ipcqnt, y = value, fill = variable),position = "dodge2") +
#   scale_fill_discrete(labels = c('linear', 'linear with fixed', 'block per capita with fixed')) +
#   theme_bw() +
#   labs( x = 'Income per equivalent adult quantiles', y = 'Changes in water bill (%)', fill = 'Tariff
#
#
# ggplot(df_sum, aes(x = ipcqnt, y = avrurban)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrurban - sdurban, ymax = avrurban + sdurban)) +
#   theme_bw() +
#   labs(y = 'Urban type', x = 'Income per equivalent adult quantiles')
#
# ggplot(df_sum, aes(x = ipcqnt, y = avrhhs)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrhhs - sdhhs, ymax = avrhhs + sdhhs)) +
#   theme_bw() +
#   labs(y = 'Household size', x = 'Income per equivalent adult quantiles')
#
# ggplot(df) +
#   geom_boxplot(aes(x = ipcqnt, y = bill_14)) +
#   theme_bw() +
#   labs(y = 'Water bill in 2014', x = 'Income per equivalent adult quantiles')
#
# ggplot(df_sum, aes(x = ipcqnt, y = rwtank)) +
#   geom_col() +
#   theme_bw() +
#   labs(y = 'Proportion with rainwater tank', x = 'Income per equivalent adult quantiles')
#
#
#
# ggplot(df) +
#   geom_boxplot(aes(x = ipcqnt, y = dif_ln)) +
#   theme_bw() +
#   labs(y = 'Difference in water bill', x = 'Income per equivalent adult quantiles')
```

```
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_ln)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income per equivalent adult quantiles') +
#   ylim(-20,20)
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_ln_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income per equivalent adult quantiles') +
#   ylim(-8,8)
#
#
#
#
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_lnwf)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income per equivalent adult quantiles') +
#   ylim(-20,20)
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_lnwf_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income per equivalent adult quantiles') +
#   ylim(-8,8)
#
#
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_brx)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income per equivalent adult quantiles') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = ipcqnt, y = dif_brx_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income per equivalent adult quantiles') +
#   ylim(-12,12)
#
#
# ## 4.2. by income -----
#
# df <- df %>% mutate(incqnt = ntile(income, 10))
# df$incqnt <- as.factor(df$incqnt)
#
#
#
# df_sum <- df %>%
#   group_by(incqnt) %>%
```

```
#   summarise(count = n(),
#             avrbill = mean(bill_14, na.rm = T),
#             sdbill = sd(bill_14, na.rm = T),
#             rwtank = sum(rwtank %in% 'yes', na.rm = T)*100/n(),
#             avrhhs = mean(hhs, na.rm = T),
#             sdhhs = sd(hhs, na.rm = T),
#             avrinc = mean(inceqa, na.rm = T),
#             sdinc = sd(income, na.rm = T),
#             avrurban = mean(urban, na.rm = T),
#             sdurban = sd(urban, na.rm = T),
#             dif_ln = mean(dif_ln),
#             dif_lnwf = mean(dif_lnwf),
#             dif_brx = mean(dif_brx))
#
#
#
# df_sum$dif_ln_prc <- df_sum$dif_ln*100/df_sum$avrbill
#
# df_sum$dif_lnwf_prc <- df_sum$dif_lnwf*100/df_sum$avrbill
#
# df_sum$dif_brx_prc <- df_sum$dif_brx*100/df_sum$avrbill
#
# plotdf <- melt(df_sum[, c('incqnt', "dif_ln_prc", "dif_lnwf_prc", "dif_brx_prc")])
#
#
#
# ggplot(plotdf) +
#   geom_col(aes(x = incqnt, y = value, fill = variable),position = "dodge2") +
#   scale_fill_discrete(labels = c('linear', 'linear with fixed', 'block per capita with fixed')) +
#   theme_bw() +
#   labs( x = 'Household income quantiles', y = 'Changes in water bill (%)', fill = 'Tariff scheme')
#
#
# ggplot(df_sum, aes(x = incqnt, y = avrurban)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrurban - sdurban, ymax = avrurban + sdurban)) +
#   theme_bw() +
#   labs(y = 'Urban type', x = 'Household income quantiles')
#
# ggplot(df_sum, aes(x = incqnt, y = avrhhs)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrhhs - sdhhs, ymax = avrhhs + sdhhs)) +
#   theme_bw() +
#   labs(y = 'Household size', x = 'Household income quantiles')
#
# ggplot(df) +
#   geom_boxplot(aes(x = incqnt, y = bill_14)) +
#   theme_bw() +
#   labs(y = 'Water bill in 2014', x = 'Household income quantiles')
#
# ggplot(df_sum, aes(x = incqnt, y = rwtank)) +
#   geom_col() +
#   theme_bw() +
```

```
#   labs(y = 'Proportion with rainwater tank', x = 'Household income quantiles')
#
#
#
# ggplot(df) +
#   geom_boxplot(aes(x = incqnt, y = dif_ln)) +
#   theme_bw() +
#   labs(y = 'Difference in water bill', x = 'Household income quantiles')
#
# ggplot(df_sum, aes(x = incqnt, y = dif_ln)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Household income quantiles') +
#   ylim(-20,20)
#
# ggplot(df_sum, aes(x = incqnt, y = dif_ln_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Household income quantiles') +
#   ylim(-8,8)
#
#
#
#
#
# ggplot(df_sum, aes(x = incqnt, y = dif_lnwf)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Household income quantiles') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = incqnt, y = dif_lnwf_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Household income quantiles') +
#   ylim(-12,12)
#
#
#
# ggplot(df_sum, aes(x = incqnt, y = dif_brx)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Household income quantiles') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = incqnt, y = dif_brx_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Household income quantiles') +
#   ylim(-12,12)
#
# ## 4.3. by income cat-----
#
```

```
#
# df$inccat <- factor(df$inccat, levels =lvls(df$inccat)[c(4,3,1,2)])
#
# df <- df[!(is.na(df$inccat)),]
#
#
# df_sum <- df %>%
#   group_by(inccat) %>%
#   summarise(count = n(),
#             avrbill = mean(bill_14, na.rm = T),
#             sdbill = sd(bill_14, na.rm = T),
#             rwtank = sum(rwtank %in% 'yes', na.rm = T)*100/n(),
#             avrhhs = mean(hhs, na.rm = T),
#             sdhhs = sd(hhs, na.rm = T),
#             avrinc = mean(inceqa, na.rm = T),
#             sdinc = sd(income, na.rm = T),
#             avrurban = mean(urban, na.rm = T),
#             sdurban = sd(urban, na.rm = T),
#             dif_ln = mean(dif_ln),
#             dif_lnwf = mean(dif_lnwf),
#             dif_brx = mean(dif_brx))
#
#
#
# df_sum$dif_ln_prc <- df_sum$dif_ln*100/df_sum$avrbill
#
# df_sum$dif_lnwf_prc <- df_sum$dif_lnwf*100/df_sum$avrbill
#
# df_sum$dif_brx_prc <- df_sum$dif_brx*100/df_sum$avrbill
#
# plotdf <- melt(df_sum[, c('inccat', "dif_ln_prc", "dif_lnwf_prc", "dif_brx_prc")])
#
#
#
# ggplot(plotdf) +
#   geom_col(aes(x = inccat, y = value, fill = variable),position = "dodge2") +
#   scale_fill_discrete(labels = c('linear', 'linear with fixed', 'block per capita with fixed')) +
#   theme_bw() +
#   labs( x ='Income categories', y = 'Changes in water bill (%)', fill = 'Tariff scheme')
#
# ggplot(df_sum, aes(x = inccat, y = avrurban)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrurban - sdurban, ymax = avrurban + sdurban)) +
#   theme_bw() +
#   labs(y = 'Urban type', x = 'Income categories')
#
# ggplot(df_sum, aes(x = inccat, y = avrhhs)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrhhs - sdhhs, ymax = avrhhs + sdhhs)) +
#   theme_bw() +
#   labs(y = 'Household size', x = 'Income categories')
#
# ggplot(df) +
```

```
#   geom_boxplot(aes(x = inccat, y = bill_14)) +
#   theme_bw() +
#   labs(y = 'Water bill in 2014', x = 'Income categories')
#
# ggplot(df_sum, aes(x = inccat, y = rwtank)) +
#   geom_col() +
#   theme_bw() +
#   labs(y = 'Proportion with rainwater tank', x = 'Income categories')
#
#
#
# ggplot(df) +
#   geom_boxplot(aes(x = inccat, y = dif_ln)) +
#   theme_bw() +
#   labs(y = 'Difference in water bill', x = 'Income categories')
#
# ggplot(df_sum, aes(x = inccat, y = dif_ln)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income categories') +
#   ylim(-20,20)
#
# ggplot(df_sum, aes(x = inccat, y = dif_ln_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income categories') +
#   ylim(-8,8)
#
#
#
#
#
#
# ggplot(df_sum, aes(x = inccat, y = dif_lnwf)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income categories') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = inccat, y = dif_lnwf_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income categories') +
#   ylim(-12,12)
#
#
#
# ggplot(df_sum, aes(x = inccat, y = dif_brx)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Income categories') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = inccat, y = dif_brx_prc)) +
```

```
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Income categories') +
#   ylim(-12,12)
#
#
# ## 4.4. by urban -----
#
#
#
# df_sum <- df %>%
#   group_by(urban) %>%
#   summarise(count = n(),
#             avrbill = mean(bill_14, na.rm = T),
#             sdbill = sd(bill_14, na.rm = T),
#             rwtank = sum(rwtank %in% 'yes', na.rm = T)*100/n(),
#             avrhhs = mean(hhs, na.rm = T),
#             sdhhs = sd(hhs, na.rm = T),
#             avrinc = mean(inceqa, na.rm = T),
#             sdinc = sd(income, na.rm = T),
#             dif_ln = mean(dif_ln),
#             dif_lnwf = mean(dif_lnwf),
#             dif_brx = mean(dif_brx))
#
#
#
# df_sum$dif_ln_prc <- df_sum$dif_ln*100/df_sum$avrbill
#
# df_sum$dif_lnwf_prc <- df_sum$dif_lnwf*100/df_sum$avrbill
#
# df_sum$dif_brx_prc <- df_sum$dif_brx*100/df_sum$avrbill
#
# plotdf <- melt(df_sum[, c('urban', "dif_ln_prc", "dif_lnwf_prc", "dif_brx_prc")])
#
#
#
# ggplot(plotdf) +
#   geom_col(aes(x = urban, y = value, fill = variable),position = "dodge2") +
#   scale_fill_discrete(labels = c('linear', 'linear with fixed', 'block per capita with fixed')) +
#   theme_bw() +
#   labs( x ='Urban density', y = 'Changes in water bill (%)', fill = 'Tariff scheme')
#
# ggplot(df_sum, aes(x = urban, y = avrhhs)) +
#   geom_point() +
#   geom_errorbar(aes(ymin = avrhhs - sdhhs, ymax = avrhhs + sdhhs)) +
#   theme_bw() +
#   labs(y = 'Household size', x = 'Urban density')
#
# ggplot(df) +
#   geom_boxplot(aes(x = urban, y = bill_14)) +
#   theme_bw() +
#   labs(y = 'Water bill in 2014', x = 'Urban density')
#
```

```
# ggplot(df_sum, aes(x = urban, y = rwtank)) +
#   geom_col() +
#   theme_bw() +
#   labs(y = 'Proportion with rainwater tank', x = 'Urban density')
#
#
#
# ggplot(df) +
#   geom_boxplot(aes(x = urban, y = dif_ln)) +
#   theme_bw() +
#   labs(y = 'Difference in water bill', x = 'Urban density')
#
# ggplot(df_sum, aes(x = urban, y = dif_ln)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Urban density') +
#   ylim(-20,20)
#
# ggplot(df_sum, aes(x = urban, y = dif_ln_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Urban density') +
#   ylim(-8,8)
#
#
#
#
#
# ggplot(df_sum, aes(x = urban, y = dif_lnwf)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Urban density') +
#   ylim(-35,35)
#
# ggplot(df_sum, aes(x = urban, y = dif_lnwf_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Urban density') +
#   ylim(-12,12)
#
#
#
# ggplot(df_sum, aes(x = urban, y = dif_brx)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (euro)', x = 'Urban density') +
#   ylim(-40,40)
#
# ggplot(df_sum, aes(x = urban, y = dif_brx_prc)) +
#   geom_col()  +
#   theme_bw() +
#   labs(y = 'Difference in water bill (%)', x = 'Urban density') +
#   ylim(-18,18)
```