Get ssb data

```
library(PxWebApiData)
library(tidyverse)
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(sf)
```

```
## Linking to GEOS 3.8.1, GDAL 3.1.4, PROJ 6.3.1
```

```
library(sp)
library(spdep)
```

```
## Loading required package: spData
```

```
library(spatialreg)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
##
## Attaching package: 'spatialreg'
```

```
## The following objects are masked from 'package:spdep':
##
##     as_dgRMatrix_listw, as_dsCMatrix_I, as_dsCMatrix_IrW,
##     as_dsTMatrix_listw, as.spam.listw, can.be.simmed, cheb_setup,
##     create_WX, do_ldet, eigen_pre_setup, eigen_setup, eigenw,
##     errorsarlm, get.ClusterOption, get.coresOption, get.mcOption,
##     get.VerboseOption, get.ZeroPolicyOption, GMargminImage, GMerrorsar,
##     griffith_sone, gstsls, Hausman.test, impacts, intImpacts,
##     Jacobian_W, jacobianSetup, l_max, lagmess, lagsarlm, lextrB,
##     lextrS, lextrW, lmSLX, LU_prepermutate_setup, LU_setup,
##     Matrix_J_setup, Matrix_setup, mcdet_setup, MCMCsamp, ME, mom_calc,
##     mom_calc_int2, moments_setup, powerWeights, sacsarlm,
##     SE_classic_setup, SE_interp_setup, SE_whichMin_setup,
##     set.ClusterOption, set.coresOption, set.mcOption,
```

```
##     set.VerboseOption, set.ZeroPolicyOption, similar.listw, spam_setup,
##     spam_update_setup, SpatialFiltering, spautolm, spBreg_err,
##     spBreg_lag, spBreg_sac, stsls, subgraph_eigenw, trW
```

```
library(broom)
library(mctest)
library(huxtable)
```

```
##
## Attaching package: 'huxtable'
```

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

```
## The following object is masked from 'package:ggplot2':
##
##     theme_grey
```

```
library(tmap)
```

Read in map 2017 (unedited)

The map is converted from 432 to 426 multipolygons in Qgis (Vektor > Geometriverktøy > Samle geometrier ...). The routine with " group_by(knr) %>% summarise(geometry = st_combine(geometry))" did not work.

```
map_2017 <- st_read("map_2017_426.gpkg")
```

```
## Reading layer `kommuner_2017' from data source `/Users/agwd/Dev/met2/data/map_2017_426.gpkg' using di
## Simple feature collection with 426 features and 1 field
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -99551.21 ymin: 6426048 xmax: 1121941 ymax: 7962744
## Projected CRS: ETRS89 / UTM zone 33N
```

```
map_2017 <- map_2017 %>%
  rename(
    knr = KOMMUNENUM
  )
```

Get knr and kNavn from local file

We want the Norvegian names, hence we have to adjust the municipality names for Vetsfold 2017.

```
load("bld06_19.Rdata")

knr_and_knavn <- bld90_19 %>%
  select(knrNavn) %>%
  unique() %>%
  # separate at the space after last digit
  separate(col = knrNavn, into = c("knr", "kNavn"), sep="\\s", extra = "merge") %>%
  filter(!knr %in% c("0712", "0715", "0729")) %>%
  bind_rows(
    c(knr = "0702", kNavn = "Holmestrand"),
    c(knr = "0709", kNavn = "Larvik"),
    c(knr = "0714", kNavn = "Hof"),
```

```
    c(knr = "0722", kNavn = "Nøtterøy"),
    c(knr = "0723", kNavn = "Tjøme"),
    c(knr = "0728", kNavn = "Lardal")
  )
```

Get ssb data via api

This document is all about getting data registered on the municipality level from SSB (one trick to get a list of all these municipality tables is to search at SSB for `(K)` and then click on `Statistikkbanken` in the left column). We will use the SSB api exposed through the `PxWebApiData` R package. This API can also be used against most Nordic SSB sister organisations. Since we have a list of the relevant municipalities for 2017 (`map_2017$knr`) we can restrict our api request to just these. We are just going to use some of these variables in the following example, but the other variables serves as examples and some might be used in further analysis. There is still plenty of relevant data at SSB that it should be quite trivial to get access to using the techniques shown below.

New dataset. Use pm2 for 2017 to avoid the big changes in Trøndelag 1.1 2018.

Get avarage price per sqr. meter single family dwellings in each municipality (426)

```
pm2_2017 <- ApiData(
    urlToData = "06035",
    Region = map_2017$knr,
    Boligtype = "01",
    Tid = "2017"
)

pm2_2017 <- pm2_2017$dataset

# list with 2 elements, we want dataset
pm2_2017 <- pm2_2017 %>%
  select(-Boligtype, -Tid) %>%
  rename(
    knr = Region
  ) %>%
  pivot_wider(id_cols = knr, names_from = "ContentsCode", values_from = "value") %>%
  rename(
    pm2 = KvPris,
    num_dwelling_sales = Omsetninger
  )

# We are tidy
dim(pm2_2017)[1]
```

```
## [1] 426
```

```
# 139 missing values. Try other years for these
pm2_missing <- pm2_2017 %>%
  filter(is.na(pm2)) %>%
  select(knr)

pm2_15to18 <- ApiData(
```

```
    urlToData = "06035",
    Region = pm2_missing$knr,
    Boligtype = "01",
    Tid = c(as.character(2010:2016), "2018")
)

tmp <- pm2_15to18$dataset %>%
  select(-Boligtype) %>%
  rename(knr = Region) %>%
  pivot_wider(id_cols = knr, names_from = c("ContentsCode", "Tid"), values_from = "value") %>%
  mutate(
    pm2 = rowMeans(select(., KvPris_2010:KvPris_2018), na.rm = TRUE),
    num_dwelling_sales = rowMeans(select(., Omsetninger_2010:Omsetninger_2018), na.rm = TRUE)
  ) %>%
  select(knr, pm2, num_dwelling_sales)

# Verty crude, we just take the mean of existing values

pm2_2017 <- pm2_2017 %>%
  left_join(tmp, by = "knr") %>%
  mutate(
    pm2 = if_else(is.na(pm2.x), as.integer(pm2.y), pm2.x),
    num_dwelling_sales =  if_else(is.na(pm2.x), as.integer(num_dwelling_sales.y), num_dwelling_sales.x)
    ) %>%
  select(knr, pm2, num_dwelling_sales)

rm(tmp)
```

Population, three categories 0-19, 20-64 and 65+

```
# Kjonn = FALSE, aggregate F+M
# The F in "F0-19" DOES NOT indicate Female
bef_2017 <- ApiData(
    urlToData = "07459",
    Region = map_2017$knr,
    Alder = list("agg:TredeltGrupperingB2",
                 c("F0-19", "F20-64", "F65+")),
    Kjonn = FALSE,
    Tid = "2017"
)

# list with 2 elements, we want dataset
bef_2017 <- bef_2017$dataset
```

```
head(bef_2017, n=6)
```

```
bef_2017 <- bef_2017 %>%
  select(Region, Alder, value) %>%
  rename(
    alder = Alder,
    knr = Region,
    bef = value,
  )
```

Region	Alder	ContentsCode	Tid	value
0101	F0-19	Personer1	2017	7033
0101	F20-64	Personer1	2017	17828
0101	F65+	Personer1	2017	5929
0104	F0-19	Personer1	2017	7453
0104	F20-64	Personer1	2017	18688
0104	F65+	Personer1	2017	6266

```
bef_2017 <- bef_2017 %>%
  pivot_wider(id_cols = knr, names_from = "alder", values_from = "bef") %>%
  # rename so we do not have to escape names
  rename(
    pop_0_19 = `F0-19`,
    pop_20_64 = `F20-64`,
    pop_65p = `F65+`
    ) %>%
  # Calculate tot. pop.
  mutate(tot_pop = pop_0_19 + pop_20_64 + pop_65p)

# we are tidy
head(bef_2017, n = 5)
```

knr	pop_0_19	pop_20_64	pop_65p	tot_pop
0101	7033	17828	5929	30790
0104	7453	18688	6266	32407
0105	13227	31829	10071	55127
0106	18728	46969	14424	80121
0111	882	2533	1102	4517

```
dim(bef_2017)[1]
```

```
## [1] 426
```

Education level (percentage each level)

```
# Get metadata
#ApiData("http://data.ssb.no/api/v0/no/table/09429",  returnMetaFrames = TRUE)
```

```
edu_2017 <- ApiData(
    urlToData = "09429",
    # we want in percent
    ContentsCode = "PersonerProsent",
```

```
    Region = map_2017$knr,
    # Drop "00" since total, 100%, also dro "09a" Uoppgit
    Nivaa = c("01", "02a", "11", "03a", "04a"),
    Kjonn = FALSE,
    Tid = "2017"
)

edu_2017 <- tibble(
  # want knr
  knr = edu_2017$dataset$Region ,
  # want level descriptions, not code
  nivaa = edu_2017$`09429: Personer 16 år og over, etter region, nivå, statistikkvariabel og år`$nivå,
  edu_lev_percent= edu_2017$dataset$value
  )

head(edu_2017, n=7)
```

knr	nivaa	edu_lev_percent
0101	Grunnskolenivå	30.7
0101	Videregående skolenivå	39.8
0101	Fagskolenivå	2.3
0101	Universitets- og høgskolenivå, kort	21.6
0101	Universitets- og høgskolenivå, lang	5.6
0104	Grunnskolenivå	29.8
0104	Videregående skolenivå	37.6

```
#make tidy
edu_2017 <- edu_2017 %>%
   pivot_wider(
     id_cols = knr,
     names_from = "nivaa",
     values_from = "edu_lev_percent"
     ) %>%
  rename(
    edu_prim = `Grunnskolenivå`,
    edu_sec = `Videregående skolenivå`,
    edu_voc = `Fagskolenivå`,
    edu_uni_short = `Universitets- og høgskolenivå, kort`,
    edu_uni_long = `Universitets- og høgskolenivå, lang`
  )

# Tidy OK
head(edu_2017)
```

```
dim(edu_2017)[1]
```

```
## [1] 426
```

knr	edu_prim	edu_sec	edu_voc	edu_uni_short	edu_uni_long
0101	30.7	39.8	2.3	21.6	5.6
0104	29.8	37.6	2.5	21.8	8.3
0105	33.2	40.2	2.5	19.9	4.2
0106	29.8	38.6	2.6	22.6	6.3
0111	25.7	42	3.3	21.9	7.1
0118	33.8	45.4	2.5	15.6	2.7

Household income

```
# Get metadata
#ApiData("http://data.ssb.no/api/v0/no/table/07183",  returnMetaFrames = TRUE)
```

```
inc_2017 <- ApiData(
    urlToData = "07183",
    # "Hushold" would also get us total number of households in muni
    ContentsCode = c(paste("Inntekt", seq(from = 1, to = 7), sep = "")),
    Region = map_2017$knr,
    Tid = "2017"
)

head(inc_2017$`07183: Husholdninger, etter region, statistikkvariabel og år`, n=9)
```

region	statistikkvariabel	år	value
Halden (-2019)	Samlet inntekt under 150 000 kr, prosent	2017	3
Halden (-2019)	Samlet inntekt 150 000 - 249 999 kr, prosent	2017	9
Halden (-2019)	Samlet inntekt 250 000 - 349 999 kr, prosent	2017	13
Halden (-2019)	Samlet inntekt 350 000 - 449 999 kr, prosent	2017	12
Halden (-2019)	Samlet inntekt 450 000 - 549 999 kr, prosent	2017	11
Halden (-2019)	Samlet inntekt 550 000 - 749 999 kr, prosent	2017	17
Halden (-2019)	Samlet inntekt 750 000 kr og over, prosent	2017	37
Moss (-2019)	Samlet inntekt under 150 000 kr, prosent	2017	3
Moss (-2019)	Samlet inntekt 150 000 - 249 999 kr, prosent	2017	8

```
inc_2017 <- tibble(
  # want knr
  knr = inc_2017$dataset$Region ,
  # want level descriptions, not code
  inc_cat = inc_2017$`07183: Husholdninger, etter region, statistikkvariabel og år`$statistikkvariabel,
  inc_lev_percent= inc_2017$dataset$value
```

```
  )

inc_2017 <- inc_2017 %>%
  pivot_wider(
     id_cols = knr,
     names_from = "inc_cat",
     values_from = "inc_lev_percent"
     ) %>%
  rename(
    # Some problem with names here, had to do names(inc_2017) above and
    # then copy the text strings
    inc_0_150 = `Samlet inntekt under 150 000 kr, prosent`,
    inc_150_250 = `Samlet inntekt 150 000 - 249 999 kr, prosent`,
    inc_250_350 = `Samlet inntekt  250 000 - 349 999 kr, prosent`,
    inc_350_450 = `Samlet inntekt  350 000 - 449 999 kr, prosent`,
    inc_450_550 = `Samlet inntekt  450 000 - 549 999 kr, prosent`,
    inc_550_750 = `Samlet inntekt  550 000 - 749 999 kr, prosent`,
    inc_750k_p = `Samlet inntekt 750 000 kr og over, prosent`
  )
# we are tidy
head(inc_2017, n=3)
```

knr	inc_0_150	inc_150_250	inc_250_350	inc_350_450	inc_450_550	inc_550_750	inc_750k_p
0101	3	9	13	12	11	17	37
0104	3	8	13	12	11	17	36
0105	2	9	13	11	11	17	37

```
dim(inc_2017)[1]
```

```
## [1] 426
```

Household net capital (wealth)

```
wealth_2017_ssb <- ApiData(
    urlToData = "10320",
    # "Hushold" would also get us total number of households in muni
    ContentsCode = c("Hushald (prosent)"),
    Nettoform = c("12", "13", "10", "25", "31", "11", "32"),
    Region = map_2017$knr,
    Tid = "2017"
)
#head(wealth_2017$dataset, n=10)
#head(wealth_2017$`10320: Hushald (prosent), etter region, nettoformue, statistikkvariabel og år`, n=10
```

```
wealth_2017 <- wealth_2017_ssb$dataset %>%
  tibble() %>%
  select(-ContentsCode, -Tid) %>%
  rename(
    knr = Region,
    net_cap = Nettoform,
```

```
    cap_lev_percent = value) %>%
  pivot_wider(
     id_cols = knr,
     names_from = "net_cap",
     values_from = "cap_lev_percent"
     ) %>%
  rename(
    net_cap_0_250k = `12`,
    net_cap_250k_500k = `13`,
    net_cap_500k_1000k = `10`,
    net_cap_1000k_2000k = `25`,
    net_cap_2000k_3000k = `31`,
    net_cap_3000k_4000k = `11`,
    net_cap_4000k_p = `32`,
  )

# we are tidy
head(wealth_2017, n=3)
```

net_cap_250k_500k	net_cap_500k_1000k	net_cap_1000k_2000k	net_cap_2000k_3000k	net_cap_3000k_
5.1	10.1	19.6	13.4	
4.1	8	17.5	15	
4.5	9.5	19.2	15.6	

```
dim(wealth_2017)[1]
```

```
## [1] 426
```

Employed by government

```
gov_emp_2017 <- ApiData(
    urlToData = "12628",
    # total num.of gov. emp. in each muni
    StatligEnhet = "0",
    # in yearly eq.
    ContentsCode = "AvtalteArsverk",
    Region = map_2017$knr,
    Tid = "2017"
)

gov_emp_2017 <- gov_emp_2017$dataset %>%
  select(Region, value) %>%
  rename(
    knr = Region,
    gov_emp = value
  )
#we are tidy
head(gov_emp_2017, n=5)
```

knr	gov_emp
0101	997
0104	614
0105	1102
0106	740
0111	18

```
dim(gov_emp_2017)[1]
```

```
## [1] 426
```

Wholesale and retail trade sales

```
sales_per_capita_2017  <- ApiData(
    urlToData = "04776",
    Region = map_2017$knr,
    Tid = "2017"
)

sales_per_capita_2017  <- sales_per_capita_2017$dataset %>%
  select(Region, value) %>%
  # spc sales per capita
  rename(
    knr = Region,
    spc = value
  )
# we are tidy
head(sales_per_capita_2017)
```

knr	spc
0101	65919
0104	77498
0105	69540
0106	80019
0111	96079
0118	19794

```
dim(sales_per_capita_2017)[1]
```

```
## [1] 426
```

Household income before and after tax in decile (percent)

Percent relative to national decile.

```
inc_dec_2017 <- ApiData(
    urlToData = "12558",
    ContentsCode = "AndelHush",
    Desiler = c(paste(0, 1:9, sep=""), "10"),
    Region = map_2017$knr,
    Tid = "2017"
)

inc_dec_2017 <- tibble(
  # want knr
  knr = inc_dec_2017$dataset$Region ,
  # want level descriptions, not code
  tax = inc_dec_2017$`12558: Husholdninger, etter region, inntekt før/etter skatt, desil, statistikkvari
  decile = inc_dec_2017$dataset$Desiler,
  percent = inc_dec_2017$dataset$value
  )

inc_dec_2017 <- inc_dec_2017 %>%
  mutate(
    tax = case_when(
      # income before and after tax
      tax == "Samlet inntekt" ~ "inc_bt",
      tax == "Inntekt etter skatt" ~ "inc_at",
      TRUE ~ as.character(tax)
    )
  ) %>%
  pivot_wider(
     id_cols = knr,
     names_from = c("tax", "decile"),
     values_from = "percent"
     )

# we are tidy
head(inc_dec_2017, n=2)
```

ot__06	inc__bt__07	inc__bt__08	inc__bt__09	inc__bt__10	inc__at__01	inc__at__02	inc__at__03	inc__at__04
10.5	10	9.7	8.3	5.7	11.5	12	10.7	10
10.3	9.1	8.6	8.4	7.5	10.5	12.2	11.6	10.7

```
dim(inc_dec_2017)[1]
```

```
## [1] 426
```

Industries

```
num_emp_2017_ssb <- ApiData(
    urlToData = "07196",
    OrgFormer = c("99","02","03","04","10","24"),
```

```
    AntAnsatte = c("99", "00", "01","02","03","04","05","15","16"),
    NACE2007 = c("01-99"),
    Region = map_2017$knr,
    Tid = "2017"
)
#so that we keep the ssb data
num_emp_2017 <- num_emp_2017_ssb$`07196: Foretak (1.1.) unntatt offentlig forvaltning og primærnæringene
num_emp_2017$knr <- num_emp_2017_ssb$dataset$Region

num_emp_2017 <- num_emp_2017 %>%
  # drop two variables
  select(-region, -statistikkvariabel, -år, -`næring (SN2007)`) %>%
  # want knr to the left, really not necessary
  select(knr,everything()) %>%
  # rename before pivot to names liked by R
  rename(
    org_form = organisasjonsform,
    num_emp_group = `antall ansatte`,
    firms = value
  )

num_emp_2017 <-  num_emp_2017 %>%
  pivot_wider(
     id_cols = knr,
     names_from = c("org_form", "num_emp_group"),
     values_from = "firms"
     )
# we are tidy
head(num_emp_2017)
```

tte	Allment aksjeselskap (ASA)_5-9 ansatte	Allment aksjeselskap (ASA)_10-19 ansatte	Allment aksjeselskap (A
0	0	0	
0	0	0	
0	0	0	
0	0	0	
0	0	0	
0	0	0	

```
dim(num_emp_2017)[1]
```

```
## [1] 426
```

Centrality index

From SSB, for the year 2018, but we use it for 2017. Probably very small changes from 2017 to 2018. Changed knr for Trøndelag to the old numbers. Split 5054 Indre Fosen into 1624 Rissa and 1718 Leksvik. The last two was given the same centrality as the combined entity.

```
centrality_2018_ed <- read_excel("Sentralitet 2018_ed.xlsx")
#View(centrality_2018_ed)
```

Put it all together

```
names(pm2_2017)
```

```
## [1] "knr"                "pm2"                "num_dwelling_sales"
```

```
data_2017 <- knr_and_knavn %>%
  left_join(pm2_2017) %>%
  left_join(bef_2017) %>%
  left_join(edu_2017) %>%
  left_join(inc_2017) %>%
  left_join(gov_emp_2017) %>%
  left_join(sales_per_capita_2017) %>%
  left_join(inc_dec_2017) %>%
  left_join(wealth_2017) %>%
  left_join(num_emp_2017) %>%
  left_join(centrality_2018_ed ) %>%
  mutate(
    log_pm2 = log(pm2),
    spc_100k = spc/100000,
    gov_emp_1k = gov_emp/1000,
    sen_ind_1k = sen_ind/1000
  ) %>%
  # rearrange
  select(knr, kNavn, pm2, log_pm2, everything())
```

```
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
## Joining, by = "knr"
```

```
#View(data_2017[, c("knr", "sen_ind")])
```

```
#head(data_2017, n=5)
dim(data_2017)
```

```
## [1] 426 109
```

We have 109 variables and 426 observations.

Add data to our map

```
# join map and data
map_data_2017 <- map_2017 %>%
  left_join(data_2017)
```

```
## Joining, by = "knr"
```

```
# write out new map with data
# if one add new variables it might be necessary to delete the file before writing new version
st_write(obj = map_data_2017, dsn = "map_data_2017.gpkg", driver = "GPKG", append=FALSE)
```

```
## Deleting layer ‘map_data_2017’ using driver ‘GPKG’
## Writing layer ‘map_data_2017’ to data source ‘map_data_2017.gpkg’ using driver ‘GPKG’
## Writing 426 features with 109 fields and geometry type Multi Polygon.
```

```
#siste
```