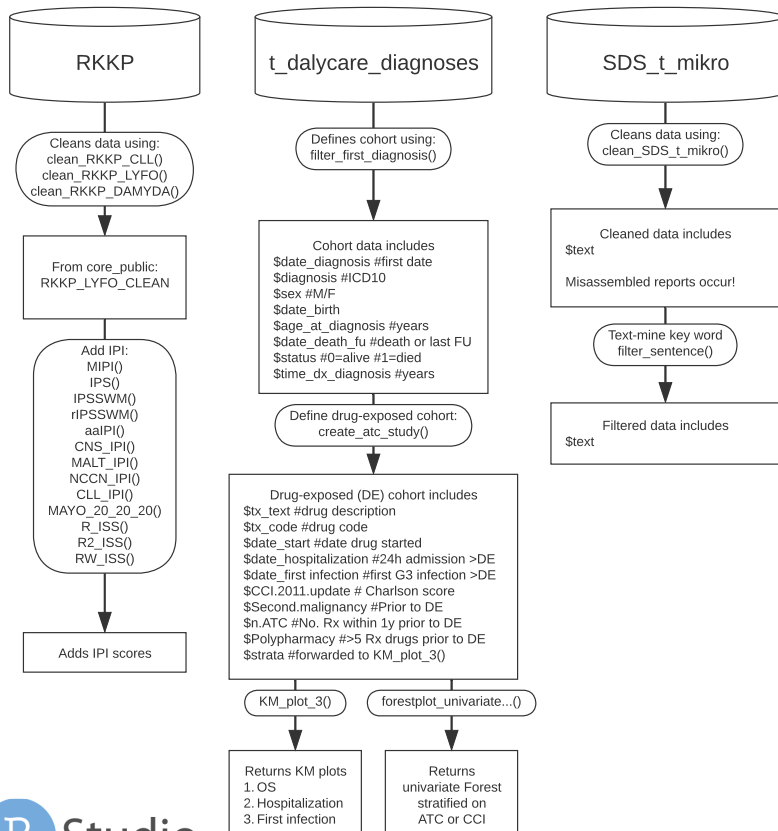# dalycare :: CHEAT SHEET

dalycare functions work with pipes (%>%) similar to dplyr

Examples

```
source('/ngc/projects2/dalyca_r/clean_r/load_dalycare_package.R') #load dalycare package
help_info() # Examples of how to load dataset(s)

#RKKP (CLL, LYFO, DaMyDa)
load_dataset('RKKP_LYFO') #load dataspecifically or
load_dataset(RKKP_DATASETS) # generally to load all three

LYFO_clean = RKKP_LYFO %>% clean_RKKP_LYFO() #clean LYFO
KM_plot(survfit(Surv(time_to_death/365, dead) ~ sex, LYFO_clean %>% filter(subtype =='DLBCL'))) #OS
MCL = RKKP_LYFO %>% clean_RKKP_LYFO() %>% MIPI() #calculate MIPI risk
KM_plot(survfit(Surv(time_to_death/365, dead) ~ MIPI, MCL)) #OS

#t_dalycare_diagnoses
load_dataset(c('patient', 't_dalycare_diagnoses')) #Loads base tables
load_dalycare_icd10() # Loads ICD10 diagnoses as a set of vectors with prefix "ICD10."
CLL = t_dalycare_diagnoses %>% filter_first_diagnosis(ICD10.CLL)
FL = t_dalycare_diagnoses %>%
    filter(tablename =='t_pato') %>% #filter biopsy-verified
    filter_first_diagnosis(ICD10.FL)
KM_plot(survfit(Surv(time_dx_death, status) ~ sex, CLL)) #OS

#create_antineoplastic_study
MM = t_dalycare_diagnoses %>%
    filter_first_diagnosis(ICD10.MM) %>%
    filter(age_at_diagnosis <50) %>%
    create_antineoplastic_study() #from date_diagnosis
MM %>%
    mutate(strata = sex) %>%
    KM_plot_3() # creates OS, hospitalization, and first infection plots
load_ATC_antineoplastics() # Loads ATC codes as a set of vectors with prefix "ATC."
CLL = t_dalycare_diagnoses %>%
    filter_first_diagnosis(ICD10.CLL) %>%
    create_antineoplastic_study(ATC.venetoclax) #Calculations/baseline from  drug-exposure
CLL %>% forest_plot_univariate_infection() #Plots univariate Forest plot default ATC 3rd level
CLL %>% forest_plot_univariate_infection('CCI') #Plots univariate Forest plot stratified on CCI category

CLL = t_dalycare_diagnoses %>%
    filter_first_diagnosis(ICD10.CLL) %>%
    create_antineoplastic_study(ATC.venetoclax, ATC.ibrutinib) #Double drug-exposure, whichever came last
CLL %>% forest_plot_univariate_os()

#SDS_t_mikro
load_dataset('SDS_t_mikro')
SDS_t_mikro %>%
    clean_SDS_t_mikro() %>%
    sample_frac(0.001) %>% #For training purposes
    filter_sentence(text, c('amyloid', 'congo')) #pattern1 or pattern2 #expect latency

# Evaluate bias
LYFO_clean %>% nrow_npatients() #Evaluate no. row and no. patients
LYFO_clean$subtype %>% table() # Evaluate diagnosis
LYFO_clean$time_to_death %>% summary() # Evaluate negative follow-up time
CLL$diagnosis %>% table() # Evaluate diagnosis
LYFO_clean %>% ggplot() + geom_histogram(aes(date_diagnosis)) #Evaluate time bias
```

RKKP

Cleans data using:
clean_RKKP_CLL()
clean_RKKP_LYFO()
clean_RKKP_DAMYDA()

↓

From core_public:
RKKP_LYFO_CLEAN

↓

Add IPI:
MIPI()
IPS()
IPSSWM()
rIPSSWM()
aaIPI()
CNS_IPI()
MALT_IPI()
NCCN_IPI()
CLL_IPI()
MAYO_20_20_20()
R_ISS()
R2_ISS()
RW_ISS()

↓

Adds IPI scores

t_dalycare_diagnoses

Defines cohort using:
filter_first_diagnosis()

↓

Cohort data includes
$date_diagnosis #first date
$diagnosis #ICD10
$sex #M/F
$date_birth
$age_at_diagnosis #years
$date_death_fu #death or last FU
$status #0=alive #1=died
$time_dx_diagnosis #years

↓

Define drug-exposed cohort:
create_atc_study()

↓

Drug-exposed (DE) cohort includes
$tx_text #drug description
$tx_code #drug code
$date_start #date drug started
$date_hospitalization #24h admission >DE
$date_first infection #first G3 infection >DE
$CCI.2011.update # Charlson score
$Second.malignancy #Prior to DE
$n.ATC #No. Rx within 1y prior to DE
$Polypharmacy #>5 Rx drugs prior to DE
$strata #forwarded to KM_plot_3()

↓

KM_plot_3() | forestplot_univariate...()

↓

Returns KM plots
1. OS
2. Hospitalization
3. First infection

Returns univariate Forest stratified on ATC or CCI

SDS_t_mikro

Cleans data using:
clean_SDS_t_mikro()

↓

Cleaned data includes
$text

Misassembled reports occur!

↓

Text-mine key word
filter_sentence()

↓

Filtered data includes
$text