

# A correlation analysis of clinical variables of TCGA-KIRC patients

This project contains a pipeline of clinical analysis of the Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) data of patients, from Genomic Data Commons Data Portal and cBioPortal.

Previously, we presented an exploratory preprocessing analysis. In this section, Chi-squared test is applied to compare two or more proportions of categorical variables and T-student test to compare the means of numeric ones regarding the levels of 'Overall\_Survival\_Status'. The Hypothesis test is performed and p-value indicates the strength of evidence in supporting the null hypothesis.

```
# Set the packages of interest
packages = c("tidyverse", "skimr", "finalfit", "rstatix", "ggpubr", "GGally", "plotly")

# if a package is installed, it will be loaded
# otherwise, the missing package(s) will be installed and loaded
package.check <- lapply(packages, FUN = function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = TRUE)
    library(x, character.only = TRUE)
  }
})

suppressMessages(library("tidyverse"))
rm(packages)
setwd(".")
```

## 1. Importing data

```
kirc_clin <- read_csv("data/kirc_clin.csv")
```

## 2. Taming data

```
kirc_clin <- kirc_clin %>%
  mutate_if(is.character, as.factor) %>%
  mutate(patient_id = as.character(patient_id),
         age = as.integer(age),
         year_diagnose = as.integer(year_diagnose))

# check
glimpse(kirc_clin)

## Rows: 537
## Columns: 29
## $ patient_id      <chr> "TCGA-3Z-A93Z", "TCGA-6D-AA2E", "TCGA-A3-3306",...
## $ age             <int> 69, 68, 67, 66, 77, 57, 59, 57, 67, 70, 52, 51,...
## $ metastasis_stg  <fct> M0, MX, M0, M0, M0, M0, M0, M0, M0, M0, M0, ...
## $ lymph_stg       <fct> N0, NX, N0, N0, N0, NX, N0, NX, N0, NX, NX, NX, ...
```

```
## $ neoplasm_stg      <fct> Stage I, Stage I, Stage I, Stage III, Stage III...
## $ tumor_stg        <fct> T1, T1, T1, T3, T3, T1, T1, T2, T2, T1, T1, T1,...
## $ disease_free_mth <dbl> 12.65, 11.89, 36.79, 47.17, 0.53, NA, NA, 49.05...
## $ disease_free_stt <fct> DiseaseFree, DiseaseFree, DiseaseFree, DiseaseF...
## $ ethnicity        <fct> not hispanic/latino, not hispanic/latino, not h...
## $ frac_genome_alter <dbl> 0.2240, 0.0000, 0.2316, 0.0494, 0.0844, 0.0883,...
## $ histology_grd     <fct> G2, G2, G3, G3, G2, G2, G3, G3, G2, G2, G1, G2,...
## $ hemoglobin        <fct> Normal, NA, NA, NA, Normal, Low, Low, NA, NA, N...
## $ neoadj_therapy    <fct> No, No, No, No, No, No, No, No, No, No, No, No,...
## $ prior_cancer      <fct> No, No, No, Yes, No, No, No, No, No, No, No, No...
## $ year_diagnose     <int> 2013, 2013, 2005, 2005, 2006, 2005, 2005, 2005,...
## $ tumor_lateral     <fct> Right, Right, Left, Right, Right, Right, Right,...
## $ long_dim          <dbl> NA, NA, 0.9, 1.0, 1.2, 1.7, 2.0, 1.7, 1.5, 1.5,...
## $ mutation_cnt      <dbl> NA, NA, NA, NA, 514, 56, 656, 577, 62, 537, 477...
## $ over_surv_mth     <dbl> 12.65, 11.89, 36.79, 47.17, 0.53, 39.13, 24.15,...
## $ over_surv_stt     <fct> LIVING, LIVING, LIVING, LIVING, LIVING, DECEASE...
## $ platelet          <fct> Normal, NA, NA, NA, Normal, NA, NA, NA, NA, NA,...
## $ race              <fct> Black/African.american, Black/African.american,...
## $ serum_ca          <fct> Normal, NA, NA, NA, Normal, NA, NA, NA, NA, NA,...
## $ gender            <fct> Male, Female, Male, Male, Female, Male, Male, M...
## $ short_dim         <dbl> NA, NA, 0.4, 0.6, 0.4, 0.4, 0.8, 0.8, 0.3, 0.4,...
## $ second_long_dim   <dbl> NA, NA, 0.7, 0.8, 1.0, 1.0, 0.9, 0.8, 1.0, 1.1,...
## $ tissue_site       <fct> OTHERS, OTHERS, A, A, A, A, A, A, A, A, A, A...
## $ person_neoplasm_stt <fct> Tumor.Free, Tumor.Free, NA, With.Tumor, Tumor.F...
## $ wbc               <fct> Normal, NA, NA, NA, Normal, NA, NA, NA, NA, NA,...
```

### 3.The dependent variable

```
# Check the number of levels. If greater than 2, it has to run a simple logistic regression presents o
table(kirc_clin$over_surv_stt, exclude = NULL)
```

```
##
## DECEASED    LIVING
##      177      360
```

### 4. Numeric variables vs. over\_surv\_stt

Correlation matrix - graphic visualization

```
cols_numeric <- kirc_clin %>%
  select_if(is.numeric) %>%
  names

kirc_clin_numeric <- kirc_clin %>%
  select(one_of(c(cols_numeric, "over_surv_stt")))

levels(kirc_clin_numeric$over_surv_stt) <- c("D","L")

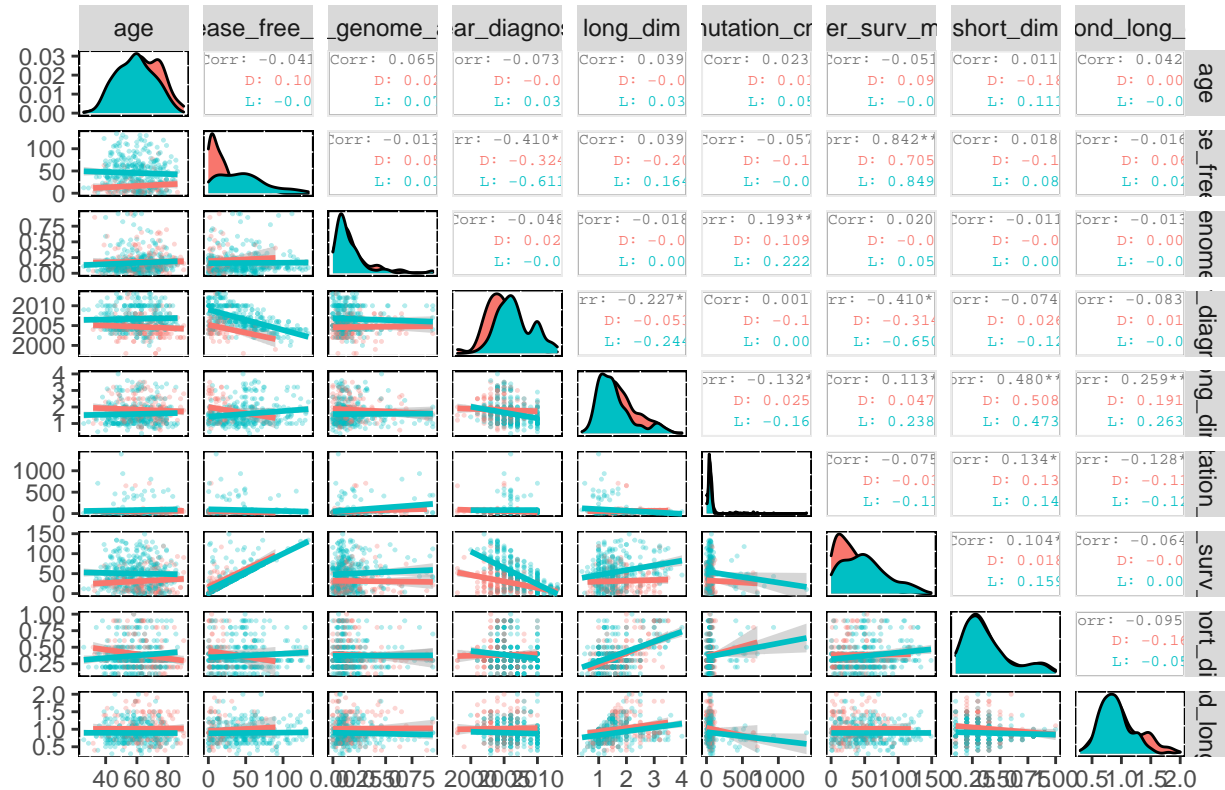
ggpairs(kirc_clin_numeric, columns = cols_numeric,
  title="Correlation matrix",
  mapping= aes(colour = over_surv_stt),
  upper = list(combo = wrap("box_no_facet", alpha=0.1),
```

```

continuous = wrap("cor", size = 2, alignPercent = 0.8)),
lower = list(continuous = wrap("smooth", alpha = 0.3, size=0.2) )) +
theme(panel.background = element_rect(color = "black", size=0.5, fill="white"),
panel.grid.major = element_blank())

```

## Correlation matrix



Run multiple T-tests on over\_surv\_stt

Transform the data into long format

```

# Put all variables in the same column except `over_surv_stt`, the grouping variable

levels(kirc_clin_numeric$over_surv_stt) <- c("DECEASED", "LIVING")

# Convert to Tidyverse
kirc_clin_numeric.long <- kirc_clin_numeric %>%
  pivot_longer(-over_surv_stt, names_to = "variables", values_to = "value")
kirc_clin_numeric.long <- kirc_clin_numeric.long[!is.na(kirc_clin_numeric.long$value), ]
kirc_clin_numeric.long$value.log <- log2(kirc_clin_numeric.long$value+1)

# OR
# kirc_clin_numeric.long <- kirc_clin_numeric %>%
#   gather(key = 'variables', value = 'value', -over_surv_stt, na.rm = TRUE) %>%
#   mutate(value.log = log2(kirc_clin_numeric.long$value+1))

kirc_clin_numeric.long %>% sample_n(6) %>% knitr::kable(.)

```

over_surv_stt	variables	value	value.log
DECEASED	long_dim	1.7	1.4329594

over_surv_stt	variables	value	value.log
LIVING	year_diagnose	2007.0	10.9715436
DECEASED	second_long_dim	1.5	1.3219281
DECEASED	year_diagnose	2002.0	10.9679467
LIVING	short_dim	0.3	0.3785116
LIVING	mutation_cnt	17.0	4.1699250

Group the data by variables and compare over\_surv\_stt groups

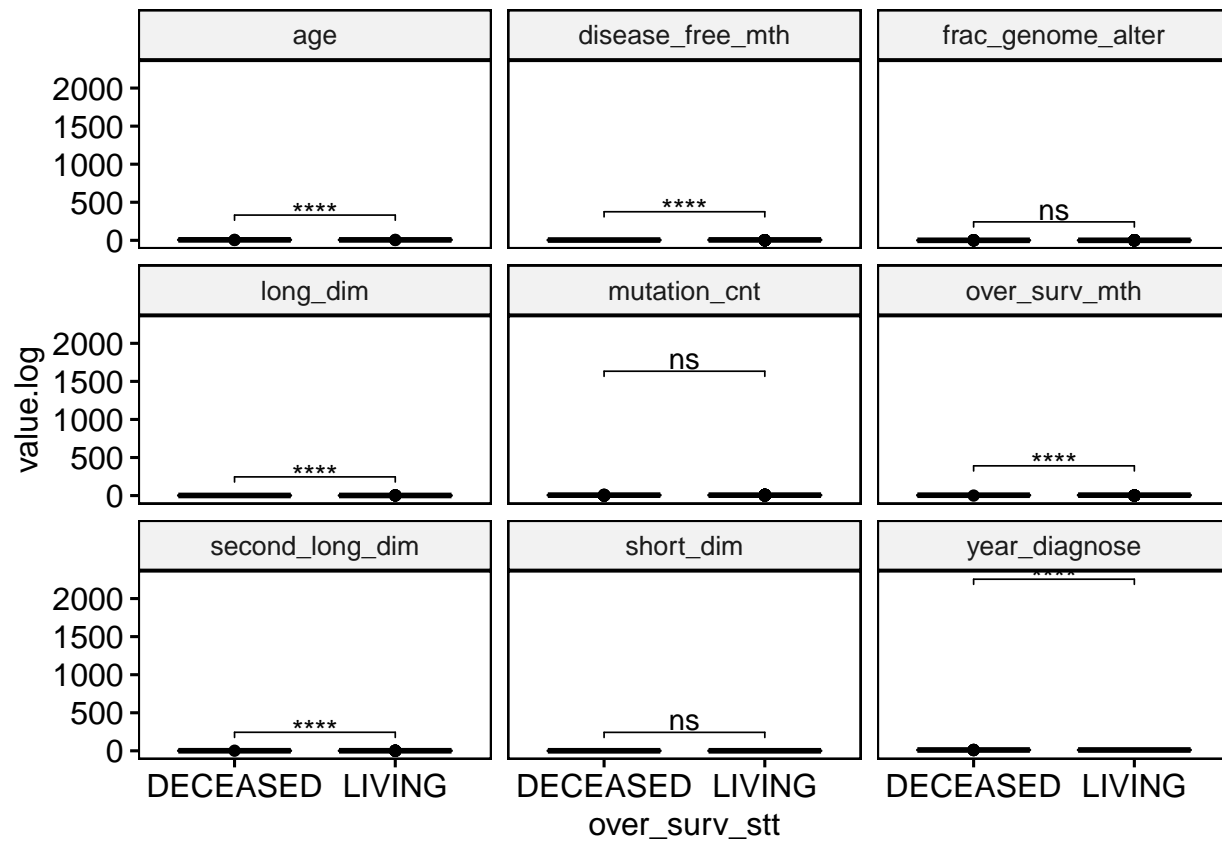
Adjust the p-values and add significance levels

```
stat.test <- kirc_clin_numeric.long %>%
  group_by(variables) %>%
  t_test(value ~ over_surv_stt) %>%
  adjust_pvalue(method = "BH") %>%
  add_significance()
stat.test %>% knitr::kable(.)
```

variables	.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
age	value	DECEASED	LIVING	177	360	4.8869626	348.1670	1.60e-06	0.0000035	****
disease_free_mth	value	DECEASED	LIVING	78	360	-10.9719698	191.6452	0.00e+000	0.0000000	****
frac_genome_alt	value	DECEASED	LIVING	175	353	1.1960301	343.6013	2.33e-01	0.2621250	ns
long_dim	value	DECEASED	LIVING	173	329	4.1297313	363.4507	4.51e-05	0.0000812	****
mutation_cnt	value	DECEASED	LIVING	153	298	-1.8305701	428.8339	6.79e-02	0.0873000	ns
over_surv_mth	value	DECEASED	LIVING	177	360	-7.3171956	450.9227	0.00e+000	0.0000000	****
second_long_dim	value	DECEASED	LIVING	173	329	4.0638742	287.9183	6.23e-05	0.0000934	****
short_dim	value	DECEASED	LIVING	173	329	0.4784069	344.6781	6.33e-01	0.6330000	ns
year_diagnose	value	DECEASED	LIVING	177	360	-8.8979681	377.0943	0.00e+000	0.0000000	****

```
# Create the plot on logscale
myplot <- ggboxplot(
  kirc_clin_numeric.long, x = "over_surv_stt", y = "value.log",
  fill = "over_surv_stt", palette = "npg", legend = "none",
  ggtheme = theme_pubr(border = TRUE)
) +
  facet_wrap(~variables)

# Add statistical test p-values
# OBS: different p-values over vaule vs. log.value!!
stat.test <- stat.test %>% add_xy_position(x = "over_surv_stt")
myplot + stat_pvalue_manual(stat.test, label = "p.adj.signif")
```



```
# Group the data by variables and do a graph for each variable
graphs <- kirc_clin_numeric.long %>%
  group_by(variables) %>%
  doo(
    ~ggboxplot(
      data = ., x = "over_surv_stt", y = "value",
      fill = "over_surv_stt", palette = "npg", legend = "none",
      ggtheme = theme_pubr()
    ) +
    geom_jitter(width = 0.05, alpha = 0.2, color = "orange"),
    result = "plots"
  )
graphs %>% knitr::kable(.)
```

[illegible]

[illegible]

[illegible]



variables

[illegible]

[illegible]

variables

[illegible]

[illegible]

[illegible]

[illegible]

```
# Add statistical tests to each corresponding plot
# variables <- graphs$variables
# for(i in 1:length(variables)){
#   graph.i <- graphs$plots[[i]] +
#     labs(title = variables[i]) +
#     #stat_pvalue_manual(stat.test[i, ], label = "p.adj.signif")
#   print(graph.i)
# }
# Error in print(graph.i) : objeto 'graph.i' não encontrado

# ggplot(kirc_clin, aes(age, fill= over_surv_stt)) +
#   geom_histogram(bins = 15, position = "dodge")
# t.test(kirc_clin$age ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(year_diagnose, fill= over_surv_stt)) +
#   geom_histogram(bins = 15, position = "dodge")
# t.test(kirc_clin$year_diagnose ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=disease_free_mth)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$disease_free_mth ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=frac_genome_alter)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$frac_genome_alter ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=long_dim)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$long_dim ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=mutation_cnt)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$mutation_cnt ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=over_surv_mth)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$over_surv_mth ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=short_dim)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
# t.test(kirc_clin$short_dim ~ kirc_clin$over_surv_stt)
#
# ggplot(kirc_clin, aes(x=over_surv_stt, y=second_long_dim)) +
#   geom_boxplot(width = .5) +
#   geom_jitter(width = 0.05, alpha = 0.2, color = "orange")
```

```
# t.test(kirc_clin$second_long_dim ~ kirc_clin$over_surv_stt)
```

Summary for continuous explanatory variables use a parametric or non-parametric test??

```
explanatory_num <- kirc_clin %>%
  select(-over_surv_stt) %>%
  select_if(is.numeric) %>%
  names
dependent <- 'over_surv_stt'

table_num <- kirc_clin %>%
  summary_factorlist(dependent, explanatory_num, p=TRUE,
    add_dependent_label=TRUE, na_include = TRUE)

knitr::kable(table_num, row.names=FALSE, align=c("l", "l", "r", "r", "r"))
```

Dependent: over_surv_stt		DECEASED	LIVING	p
age	Mean (SD)	64.2 (12.0)	58.8 (11.9)	<0.001
disease_free_mth	Mean (SD)	16.4 (18.2)	45.5 (31.5)	<0.001
frac_genome_alter	Mean (SD)	0.2 (0.2)	0.2 (0.2)	0.230
year_diagnose	Mean (SD)	2004.6 (2.4)	2006.7 (2.7)	<0.001
long_dim	Mean (SD)	1.8 (0.6)	1.6 (0.7)	<0.001
mutation_cnt	Mean (SD)	60.5 (90.9)	80.7 (142.7)	0.111
over_surv_mth	Mean (SD)	31.6 (25.1)	50.5 (33.6)	<0.001
short_dim	Mean (SD)	0.4 (0.2)	0.4 (0.2)	0.631
second_long_dim	Mean (SD)	1.0 (0.3)	0.9 (0.3)	<0.001

```
# Correlation Matrix
# Pearson's (normal distribution) or Spearman (not-normal) correlations
corr_num <- kirc_clin %>%
  select_if(is.numeric) %>%
  drop_na()

# Check the correlation between variables to exclude the highly correlated
cor_matrix <- cor(corr_num, method = "spearman")
cor_matrix <- round(cor_matrix, 2)

cor_matrix
```

```
##           age disease_free_mth frac_genome_alter year_diagnose
## age           1.00          -0.06             0.17          -0.03
## disease_free_mth -0.06           1.00          -0.05          -0.32
## frac_genome_alter 0.17          -0.05             1.00          -0.04
## year_diagnose    -0.03          -0.32          -0.04           1.00
## long_dim          0.03           0.02             0.05          -0.29
## mutation_cnt      0.26          -0.03             0.23           0.04
## over_surv_mth     -0.02           0.83          -0.02          -0.45
## short_dim         0.06           0.06          -0.03          -0.19
## second_long_dim   0.02          -0.02          -0.02          -0.06
##
## long_dim mutation_cnt over_surv_mth short_dim second_long_dim
## age           0.03           0.26          -0.02           0.06           0.02
## disease_free_mth 0.02          -0.03           0.83           0.06          -0.02
## frac_genome_alter 0.05           0.23          -0.02          -0.03          -0.02
```



```
## year_diagnose      -0.29      0.04      -0.45      -0.19      -0.06
## long_dim           1.00     -0.05      0.14      0.37      0.34
## mutation_cnt      -0.05      1.00     -0.02      0.11     -0.10
## over_surv_mth       0.14     -0.02      1.00      0.14     -0.01
## short_dim          0.37      0.11      0.14      1.00     -0.07
## second_long_dim     0.34     -0.10     -0.01     -0.07      1.00
```

## 5. Categorical variables vs. over\_surv\_stt

Tabulation and chi-square test

```
# t_metastasis_stg <- table(kirc_clin$metastasis_stg, kirc_clin$over_surv_stt, exclude = NULL)
# t_metastasis_stg <- addmargins(round(100*prop.table(t_metastasis_stg)))
# t_metastasis_stg
# chisq.test(x = kirc_clin$metastasis_stg, y = kirc_clin$over_surv_stt)
```

Summary for categorical explanatory variables Chi-squared warnings will be generated when the expected count in any cell is less than 5.

```
explanatory_char <- kirc_clin %>%
  select(-over_surv_stt) %>%
  select_if(is.factor) %>%
  names

dependent <- 'over_surv_stt'

table_char <- kirc_clin %>%
  summary_factorlist(dependent, explanatory_char, p=TRUE,
    add_dependent_label=TRUE, na_include = TRUE)

## Warning in chisq.test(tumor_stg, over_surv_stt): Chi-squared approximation may
## be incorrect

## Warning in chisq.test(histology_grd, over_surv_stt): Chi-squared approximation
## may be incorrect

## Warning in chisq.test(hemoglobin, over_surv_stt): Chi-squared approximation may
## be incorrect

## Warning in chisq.test(tumor_lateral, over_surv_stt): Chi-squared approximation
## may be incorrect

## Warning in chisq.test(race, over_surv_stt): Chi-squared approximation may be
## incorrect

## Warning in chisq.test(serum_ca, over_surv_stt): Chi-squared approximation may be
## incorrect

## Warning in chisq.test(wbc, over_surv_stt): Chi-squared approximation may be
## incorrect

knitr::kable(table_char, row.names=FALSE, align=c("l", "l", "r", "r", "r"))
```

Dependent: over_surv_stt		DECEASED	LIVING	p
metastasis_stg	M0	110 (62.1)	316 (87.8)	<0.001
	M1	64 (36.2)	15 (4.2)	
	MX	3 (1.7)	27 (7.5)	

Dependent: over_surv_stt		DECEASED	LIVING	p
lymph_stg	(Missing)	0 (0.0)	2 (0.6)	0.001
	N0	85 (48.0)	155 (43.1)	
	N1	12 (6.8)	5 (1.4)	
	NX	80 (45.2)	200 (55.6)	
neoplasms_stg	Stage I	45 (25.4)	224 (62.2)	<0.001
	Stage II	13 (7.3)	44 (12.2)	
	Stage III	51 (28.8)	74 (20.6)	
	Stage IV	67 (37.9)	16 (4.4)	
	(Missing)	1 (0.6)	2 (0.6)	
tumor_stg	T1	50 (28.2)	225 (62.5)	<0.001
	T2	21 (11.9)	48 (13.3)	
	T3	96 (54.2)	86 (23.9)	
	T4	10 (5.6)	1 (0.3)	
disease_free_stt	DiseaseFree	0 (0.0)	311 (86.4)	<0.001
	Recurred/Progressed	78 (44.1)	49 (13.6)	
	(Missing)	99 (55.9)	0 (0.0)	
ethnicity	hispanic/latino	3 (1.7)	23 (6.4)	0.017
	not hispanic/latino	132 (74.6)	227 (63.1)	
	(Missing)	42 (23.7)	110 (30.6)	
histology_grd	G1	0 (0.0)	14 (3.9)	<0.001
	G2	45 (25.4)	185 (51.4)	
	G3	74 (41.8)	133 (36.9)	
	G4	57 (32.2)	21 (5.8)	
	GX	1 (0.6)	4 (1.1)	
	(Missing)	0 (0.0)	3 (0.8)	
hemoglobin	Elevated	3 (1.7)	2 (0.6)	<0.001
	Low	118 (66.7)	145 (40.3)	
	Normal	43 (24.3)	143 (39.7)	
	(Missing)	13 (7.3)	70 (19.4)	
neoadj_therapy	No	167 (94.4)	352 (97.8)	0.069
	Yes	10 (5.6)	8 (2.2)	
prior_cancer	No	153 (86.4)	306 (85.0)	0.753
	Yes	24 (13.6)	54 (15.0)	
tumor_lateral	Bilateral	0 (0.0)	1 (0.3)	0.023
	Left	98 (55.4)	155 (43.1)	
	Right	79 (44.6)	204 (56.7)	
platelet	Elevated	30 (16.9)	8 (2.2)	<0.001
	Low	20 (11.3)	26 (7.2)	
	Normal	108 (61.0)	252 (70.0)	
	(Missing)	19 (10.7)	74 (20.6)	
race	Asian	1 (0.6)	7 (1.9)	0.032
	Black/African.american	11 (6.2)	45 (12.5)	
	White	163 (92.1)	303 (84.2)	
	(Missing)	2 (1.1)	5 (1.4)	
serum_ca	Elevated	9 (5.1)	1 (0.3)	0.001
	Low	66 (37.3)	138 (38.3)	
	Normal	62 (35.0)	89 (24.7)	
	(Missing)	40 (22.6)	132 (36.7)	
gender	Female	65 (36.7)	126 (35.0)	0.767
	Male	112 (63.3)	234 (65.0)	
tissue_site	A	11 (6.2)	68 (18.9)	<0.001
	B	112 (63.3)	191 (53.1)	

Dependent: over_surv_stt		DECEASED	LIVING	p
person_neoplasm_stt	C	49 (27.7)	78 (21.7)	<0.001
	OTHERS	5 (2.8)	23 (6.4)	
	Tumor.Free	57 (32.2)	304 (84.4)	
	With.Tumor	105 (59.3)	36 (10.0)	
wbc	(Missing)	15 (8.5)	20 (5.6)	0.011
	Elevated	46 (26.0)	118 (32.8)	
	Low	6 (3.4)	3 (0.8)	
	Normal	104 (58.8)	164 (45.6)	
(Missing)		21 (11.9)	75 (20.8)	

```

# Dropping levels with narrow distributions -> check warnings ()
# Group some levels or drop one (NULL = 'level') when grouping is not possible

kirc_clin2 <- kirc_clin %>%
  mutate(neoplasm_stg = fct_collapse(neoplasm_stg, 'Stage I-II' = c('Stage I','Stage II'), 'Stage III-IV' = c('Stage III','Stage IV')),
         tumor_stg = fct_collapse(tumor_stg, 'T1-T2' = c('T1','T2'), 'T3-T4' = c('T3','T4')),
         histology_grd = fct_collapse(histology_grd, 'G1-G2' = c('G1','G2'), 'G3-G4' = c('G3','G4')),
         hemoglobin = fct_collapse(hemoglobin, 'Not.Low' = c('Normal', 'Elevated')),
         serum_ca = fct_collapse(serum_ca, 'Not.Low' = c('Normal', 'Elevated')),
         wbc = fct_collapse(wbc, 'Not.Elevated' = c('Low', 'Normal')))

kirc_clin2 <- kirc_clin2 %>%
  mutate(histology_grd = fct_recode(histology_grd, NULL = 'GX'),
         tumor_lateral = fct_recode(tumor_lateral, NULL = 'Bilateral'),
         race = fct_recode(race, NULL = 'Asian'))

table_char2 <- kirc_clin2 %>%
  summary_factorlist(dependent, explanatory_char, p=TRUE,
                    add_dependent_label=TRUE, na_include = TRUE)

knitr::kable(table_char2, row.names=FALSE, align=c("l", "l", "r", "r", "r"))

```

Dependent: over_surv_stt		DECEASED	LIVING	p
metastasis_stg	M0	110 (62.1)	316 (87.8)	<0.001
	M1	64 (36.2)	15 (4.2)	
	MX	3 (1.7)	27 (7.5)	
	(Missing)	0 (0.0)	2 (0.6)	
lymph_stg	N0	85 (48.0)	155 (43.1)	0.001
	N1	12 (6.8)	5 (1.4)	
	NX	80 (45.2)	200 (55.6)	
neoplasm_stg	Stage I-II	58 (32.8)	268 (74.4)	<0.001
	Stage III-IV	118 (66.7)	90 (25.0)	
	(Missing)	1 (0.6)	2 (0.6)	
tumor_stg	T1-T2	71 (40.1)	273 (75.8)	<0.001
	T3-T4	106 (59.9)	87 (24.2)	
disease_free_stt	DiseaseFree	0 (0.0)	311 (86.4)	<0.001
	Recurred/Progressed	78 (44.1)	49 (13.6)	
	(Missing)	99 (55.9)	0 (0.0)	
ethnicity	hispanic/latino	3 (1.7)	23 (6.4)	0.017
	not hispanic/latino	132 (74.6)	227 (63.1)	

Dependent: over_surv_stt		DECEASED	LIVING	p
histology_grd	(Missing)	42 (23.7)	110 (30.6)	<0.001
	G1-G2	45 (25.4)	199 (55.3)	
	G3-G4	131 (74.0)	154 (42.8)	
hemoglobin	(Missing)	1 (0.6)	7 (1.9)	<0.001
	Not.Low	46 (26.0)	145 (40.3)	
	Low	118 (66.7)	145 (40.3)	
neoadj_therapy	(Missing)	13 (7.3)	70 (19.4)	0.069
	No	167 (94.4)	352 (97.8)	
	Yes	10 (5.6)	8 (2.2)	
prior_cancer	No	153 (86.4)	306 (85.0)	0.753
	Yes	24 (13.6)	54 (15.0)	
	(Missing)	13 (7.3)	70 (19.4)	
tumor_lateral	Left	98 (55.4)	155 (43.1)	0.010
	Right	79 (44.6)	204 (56.7)	
	(Missing)	0 (0.0)	1 (0.3)	
platelet	Elevated	30 (16.9)	8 (2.2)	<0.001
	Low	20 (11.3)	26 (7.2)	
	Normal	108 (61.0)	252 (70.0)	
race	(Missing)	19 (10.7)	74 (20.6)	0.032
	Black/African.american	11 (6.2)	45 (12.5)	
	White	163 (92.1)	303 (84.2)	
serum_ca	(Missing)	3 (1.7)	12 (3.3)	0.028
	Not.Low	71 (40.1)	90 (25.0)	
	Low	66 (37.3)	138 (38.3)	
gender	(Missing)	40 (22.6)	132 (36.7)	0.767
	Female	65 (36.7)	126 (35.0)	
	Male	112 (63.3)	234 (65.0)	
tissue_site	A	11 (6.2)	68 (18.9)	<0.001
	B	112 (63.3)	191 (53.1)	
	C	49 (27.7)	78 (21.7)	
person_neoplasm_stt	OTHERS	5 (2.8)	23 (6.4)	<0.001
	Tumor.Free	57 (32.2)	304 (84.4)	
	With.Tumor	105 (59.3)	36 (10.0)	
wbc	(Missing)	15 (8.5)	20 (5.6)	0.018
	Elevated	46 (26.0)	118 (32.8)	
	Not.Elevated	110 (62.1)	167 (46.4)	
	(Missing)	21 (11.9)	75 (20.8)	

## 6. saving dataset for regression model

```
kirc_glm <- kirc_clin2
write_csv(kirc_glm, path = "data/kirc_glm.csv")
```

## Further analysis

- A logistic regression analysis of each clinical variable weight.

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
```

```

## Running under: Ubuntu 18.04.5 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblas-p-r0.2.20.so
##
## locale:
## [1] LC_CTYPE=pt_BR.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=pt_BR.UTF-8      LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=pt_BR.UTF-8  LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=pt_BR.UTF-8     LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=pt_BR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] plotly_4.9.2.1  GGally_2.0.0    ggpubr_0.4.0    rstatix_0.6.0
## [5] finalfit_1.0.2  skimr_2.1.2     forcats_0.5.0   stringr_1.4.0
## [9] dplyr_1.0.2     purrr_0.3.4     readr_1.4.0     tidyr_1.1.2
## [13] tibble_3.0.3    ggplot2_3.3.2   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] nlme_3.1-149      fs_1.5.0         lubridate_1.7.9   RColorBrewer_1.1-2
## [5] httr_1.4.2        ggsci_2.9         repr_1.1.0        tools_4.0.3
## [9] backports_1.1.10  utf8_1.1.4       R6_2.4.1          mgcv_1.8-33
## [13] DBI_1.1.0         lazyeval_0.2.2   colorspace_1.4-1  withr_2.3.0
## [17] tidyselect_1.1.0  curl_4.3         compiler_4.0.3    cli_2.0.2
## [21] rvest_0.3.6       mice_3.11.0      xml2_1.3.2        labeling_0.3
## [25] scales_1.1.1      digest_0.6.25    foreign_0.8-79    rmarkdown_2.4
## [29] rio_0.5.16        base64enc_0.1-3  pkgconfig_2.0.3   htmltools_0.5.0
## [33] highr_0.8         dbplyr_1.4.4     htmlwidgets_1.5.2  rlang_0.4.7
## [37] readxl_1.3.1      rstudioapi_0.11  farver_2.0.3      generics_0.0.2
## [41] jsonlite_1.7.1    zip_2.1.1        car_3.0-10        magrittr_1.5
## [45] Matrix_1.2-18     Rcpp_1.0.5       munsell_0.5.0     fansi_0.4.1
## [49] abind_1.4-5       lifecycle_0.2.0  stringi_1.5.3     yaml_2.2.1
## [53] carData_3.0-4     plyr_1.8.6       grid_4.0.3        blob_1.2.1
## [57] crayon_1.3.4      lattice_0.20-41  haven_2.3.1       splines_4.0.3
## [61] hms_0.5.3         knitr_1.30       pillar_1.4.6      boot_1.3-25
## [65] ggsignif_0.6.0    reprex_0.3.0     glue_1.4.2        evaluate_0.14
## [69] data.table_1.13.0 modelr_0.1.8      vctrs_0.3.4       cellranger_1.1.0
## [73] gtable_0.3.0      reshape_0.8.8    assertthat_0.2.1  xfun_0.18
## [77] openxlsx_4.2.2    broom_0.7.1      survival_3.2-7     viridisLite_0.3.0
## [81] ellipsis_0.3.1

```