

# Supplement 4

Reviewer analysis requests: Number of pain sites

Peter Kamerman

Last knitted: 16 October 2021

## Contents

<b>1</b>	<b>Import and check data</b>	<b>1</b>
<b>2</b>	<b>Data analysis</b>	<b>4</b>
2.1	Process data . . . . .	4
2.2	Summary statistics . . . . .	4
2.2.1	Total group . . . . .	4
2.2.2	For each categorical variable . . . . .	5
2.2.2.1	By diabetes . . . . .	5
2.2.2.2	By TB . . . . .	5
2.3	Exploratory plots . . . . .	6
2.3.1	For each categorical variable . . . . .	6
2.3.1.1	Count by diabetes . . . . .	6
2.3.1.2	Count by TB . . . . .	9
2.3.2	Continuous variables . . . . .	11
2.3.2.1	Beck's Depression Inventory II . . . . .	12
<b>3</b>	<b>Session information</b>	<b>13</b>

---

## 1 Import and check data

```
# Import
data <- read_rds('data-cleaned/data-pain-sites.rds')
demo <- read_rds('data-cleaned/data-demographics.rds') %>%
  select(ID, Site, Diabetes, TB, BDI) # Select required columns

# Check
## Pain sites
dim(data)

## [1] 599 21
names(data)

## [1] "ID" "Head" "Throat"
## [4] "Shoulder" "Arms" "Elbows"
## [7] "Wrists.Hands" "Chest" "Upper_back"
```

```
## [10] "Lower_back"      "Abdomen"          "Cervical_spine"
## [13] "Thoracic_spine"  "Lumbosacral_spine" "Groin"
## [16] "Hips"            "Legs"              "Knees"
## [19] "Ankles.Feet"     "Buttocks"          "Site"
```

```
glimpse(data)
```

```
## Rows: 599
## Columns: 21
## $ ID          <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB78"~
## $ Head        <chr> "No", "No", "No", "Yes", "Yes", "No", "No", "No", "Y~
## $ Throat      <chr> "No", "No", "No", "No", "No", "No", "No", "No", "Yes~
## $ Shoulder    <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Arms        <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Elbows      <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Wrists.Hands <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Chest       <chr> "No", "No", "No", "Yes", "No", "No", "No", "No", "Ye~
## $ Upper_back  <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Lower_back  <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Abdomen     <chr> "No", "No", "Yes", "Yes", "No", "No", "Yes", "No", "~
## $ Cervical_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Thoracic_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Lumbosacral_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Groin       <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ Hips        <chr> "No", "No", "No", "No", "No", "No", "No", "Yes", "No~
## $ Legs        <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "N~
## $ Knees       <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "N~
## $ Ankles.Feet <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "N~
## $ Buttocks    <chr> "No", "No", "No", "No", "No", "No", "No", "Yes", "No~
## $ Site        <chr> "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP"~
```

```
data %>%
  select(-ID, -Site) %>%
  mutate_if(is.character, factor) %>%
  skim()
```

Table 1: Data summary

Name	Piped data
Number of rows	599
Number of columns	19
Column type frequency:	
factor	19
Group variables	None

## Variable type: factor

skim_variable	missing	complete	n_unique	top_counts
Head	0	599	2	No: 402, Yes: 197
Throat	0	599	2	No: 580, Yes: 19
Shoulder	0	599	2	No: 552, Yes: 47

skim_variable	missing	complete	n_unique	top_counts
Arms	0	599	2	No: 574, Yes: 25
Elbows	0	599	2	No: 577, Yes: 22
Wrists.Hands	0	599	2	No: 563, Yes: 36
Chest	0	599	2	No: 481, Yes: 118
Upper_back	0	599	1	No: 599
Lower_back	0	599	2	No: 557, Yes: 42
Abdomen	0	599	2	No: 440, Yes: 159
Cervical_spine	0	599	2	No: 569, Yes: 30
Thoracic_spine	0	599	2	No: 525, Yes: 74
Lumbosacral_spine	0	599	2	No: 506, Yes: 93
Groin	0	599	2	No: 545, Yes: 54
Hips	0	599	2	No: 559, Yes: 40
Legs	0	599	2	No: 500, Yes: 99
Knees	0	599	2	No: 515, Yes: 84
Ankles.Feet	0	599	2	No: 415, Yes: 184
Buttocks	0	599	2	No: 580, Yes: 19

### ## Demographics

```
dim(demo)
```

```
## [1] 599 5
```

```
names(demo)
```

```
## [1] "ID" "Site" "Diabetes" "TB" "BDI"
```

```
glimpse(demo)
```

```
## Rows: 599
```

```
## Columns: 5
```

```
## $ ID <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB78", "RPB80"~
```

```
## $ Site <chr> "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "~
```

```
## $ Diabetes <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "~
```

```
## $ TB <chr> "No", "No", "No", "No", "Yes", "No", "Yes", "No", "No", "No", "~
```

```
## $ BDI <dbl> 23, 25, 5, 25, 24, 24, 8, 22, 9, 8, 8, 7, 3, 9, 26, 5, 33, NA~
```

```
demo %>%
```

```
  select(-ID, -Site) %>%
```

```
  mutate_if(is.character, factor) %>%
```

```
  skim()
```

Table 3: Data summary

Name	Piped data
Number of rows	599
Number of columns	3
Column type frequency:	
factor	2
numeric	1
Group variables	None

Variable type: factor

skim_variable	missing	complete	n_unique	top_counts
Diabetes	241	358	2	No: 342, Yes: 16
TB	228	371	2	No: 261, Yes: 110

Variable type: numeric

skim_variable	missing	complete	mean	sd	p0	p25	p50	p75	p100
BDI	348	251	17.91	9.74	1	10	17	25	46

## 2 Data analysis

### 2.1 Process data

```
# Remove upper_back (only one outcome -- no pain)
data <- data[, !(names(data) %in% c('Upper_back'))]

# Calculate number of sites per individual
pain_count <- data %>%
  # Format as long data
  pivot_longer(cols = -c(ID, Site),
               names_to = 'pain_site',
               values_to = 'pain_present') %>%
  group_by(ID) %>%
  summarise(Count = sum(pain_present == 'Yes', na.rm = TRUE))

# Join datasets
analysis_set <- left_join(pain_count, demo)
```

### 2.2 Summary statistics

#### 2.2.1 Total group

```
analysis_set %>%
  select(Count) %>%
  skim()
```

Table 6: Data summary

Name	Piped data
Number of rows	599
Number of columns	1

Column type frequency:	
numeric	1
Group variables	None

**Variable type: numeric**

skim_variable	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	0	599	2.26	1.7	0	1	2	3	12

## 2.2.2 For each categorical variable

```
analysis_set %>%
  select(Count, Diabetes) %>%
  filter(complete.cases(.)) %>%
  group_by(Diabetes) %>%
  skim()
```

### 2.2.2.1 By diabetes

Table 8: Data summary

Name	Piped data
Number of rows	358
Number of columns	2
Column type frequency:	
numeric	1
Group variables	Diabetes

**Variable type: numeric**

skim_variable	Diabetes	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	No	0	342	2.30	1.83	0	1.00	2	3	12
Count	Yes	0	16	2.88	1.63	0	1.75	3	4	5

```
analysis_set %>%
  select(Count, TB) %>%
  filter(complete.cases(.)) %>%
  group_by(TB) %>%
  skim()
```

### 2.2.2.2 By TB

Table 10: Data summary

Name	Piped data
Number of rows	371
Number of columns	2
Column type frequency: numeric	1
Group variables	TB

### Variable type: numeric

skim_variable	TB	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	No	0	261	2.28	1.90	0	1	2	3	12
Count	Yes	0	110	2.43	1.61	0	1	2	3	7

## 2.3 Exploratory plots

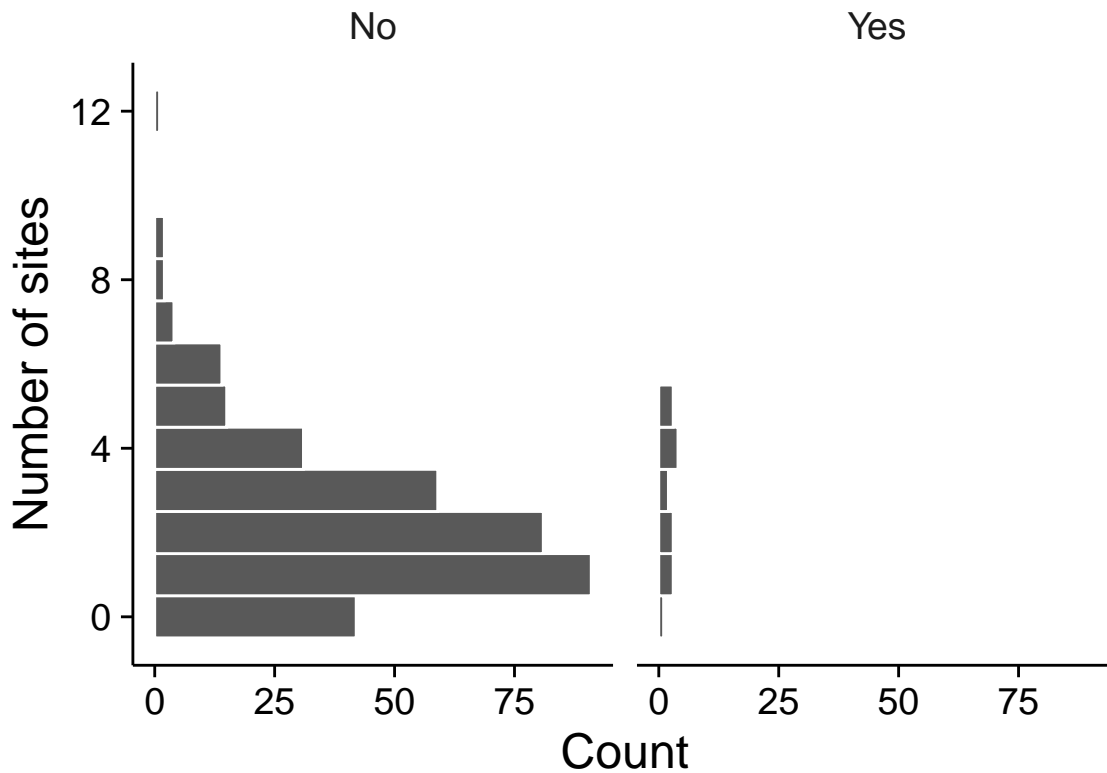
### 2.3.1 For each categorical variable

```
analysis_set %>%
  select(Count, Diabetes) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  labs(title = 'Diabetes mellitus',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                           analysis_set$Diabetes), ]}))'),
       x = 'Number of sites',
       y = 'Count') +
  facet_wrap(~Diabetes) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))
```

#### 2.3.1.1 Count by diabetes

# Diabetes mellitus

(Complete cases = 358)



```
diabetes <- analysis_set %>%
  select(Count, Diabetes) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = Diabetes) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#FFFFFF',
             colour = '#000000',
             stroke = 0.8,
             alpha = 0.5) +
  geom_boxplot(outlier.colour = '#FFFFFF',
              outlier.size = 0,
              size = 0.8,
              alpha = 0) +
  scale_x_discrete(labels = c('No', 'Yes')) +
  labs(title = 'Diabetes mellitus',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Diabetes), ]})}')),
       y = 'Number of sites') +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
```

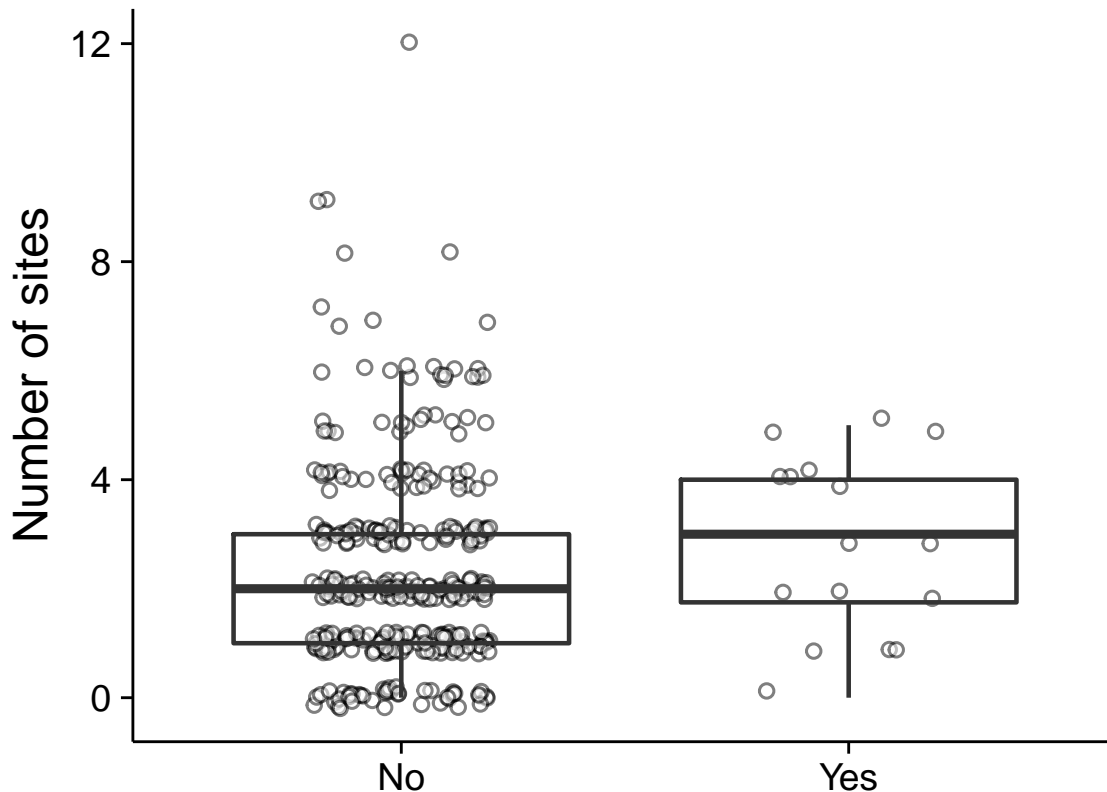
```

plot.caption = element_text(size = 12),
panel.grid = element_blank(),
axis.title.x = element_blank(),
axis.text = element_text(colour = '#000000'),
axis.line = element_line(size = 0.5),
axis.ticks = element_line(size = 0.5)); diabetes

```

## Diabetes mellitus

(Complete cases = 358)



```
wilcox.test(Count ~ Diabetes, data = analysis_set)
```

```

##
## Wilcoxon rank sum test with continuity correction
##
## data: Count by Diabetes
## W = 2092.5, p-value = 0.105
## alternative hypothesis: true location shift is not equal to 0

```

```

analysis_set %>%
  select(Count, TB) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
    colour = '#FFFFFF') +

```

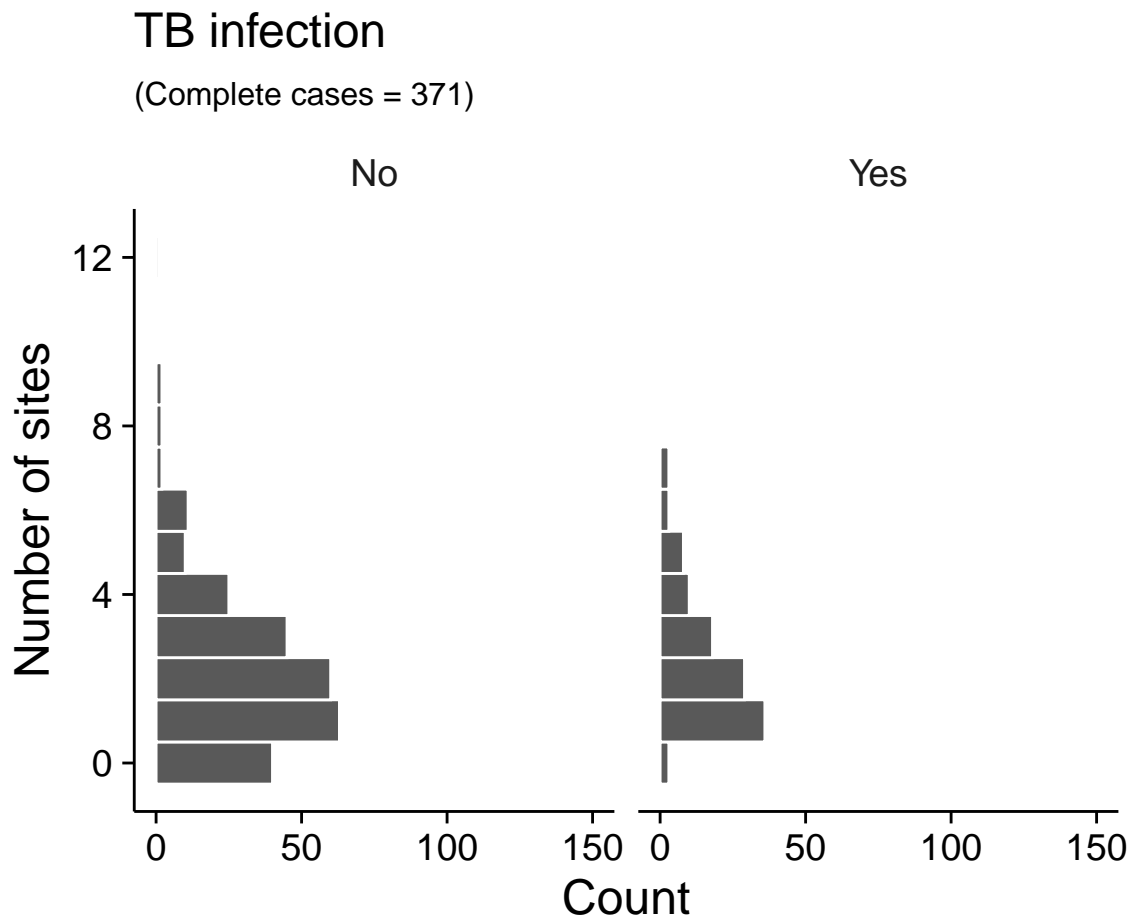


```

scale_y_continuous(limits = c(0, 150),
                   breaks = c(0, 50, 100, 150)) +
labs(title = 'TB infection',
     subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
       analysis_set$TB), ]})}')),
     x = 'Number of sites',
     y = 'Count') +
facet_wrap(~TB) +
coord_flip() +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      panel.grid = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5))

```

### 2.3.1.2 Count by TB



```

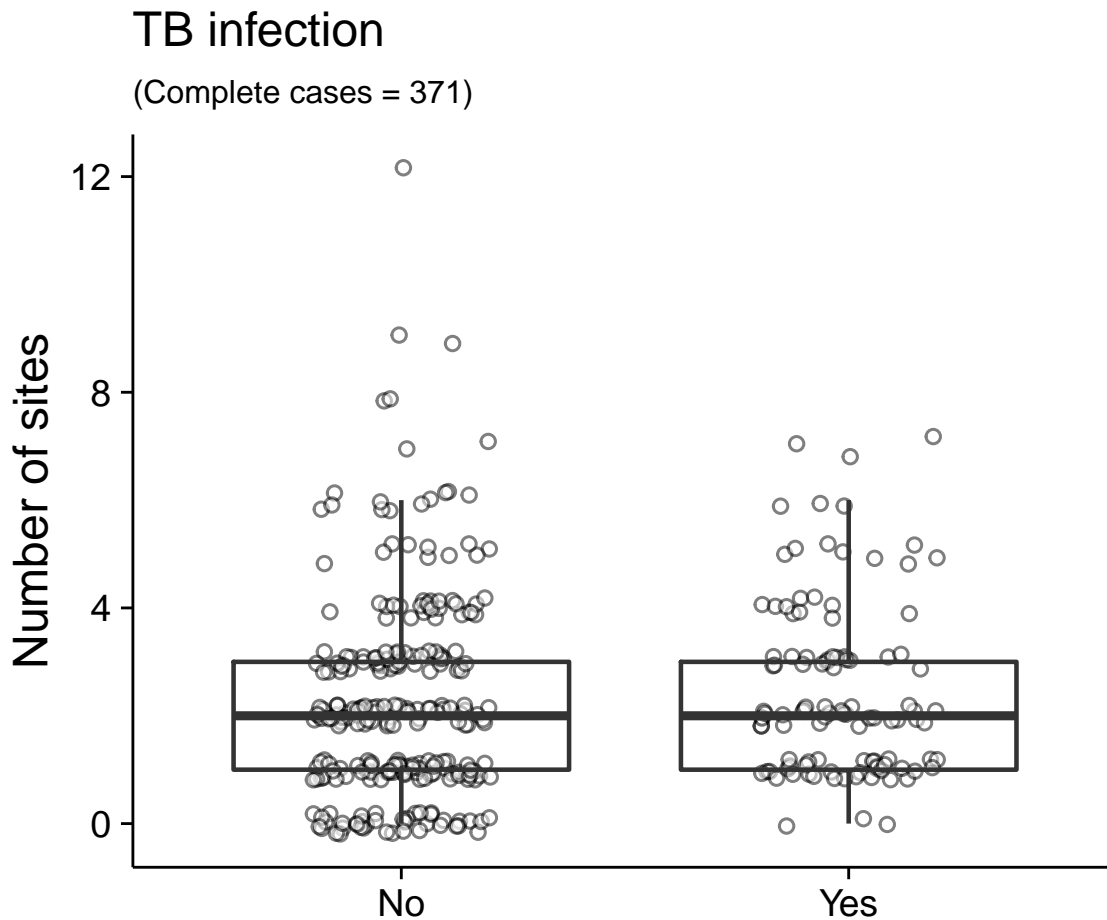
tb <- analysis_set %>%
  select(Count, TB) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = TB) +

```

```

geom_point(size = 2,
           shape = 21,
           position = position_jitter(height = 0.2, width = 0.2),
           fill = '#FFFFFF',
           colour = '#000000',
           stroke = 0.8,
           alpha = 0.5) +
geom_boxplot(outlier.colour = '#FFFFFF',
             outlier.size = 0,
             size = 0.8,
             alpha = 0) +
scale_x_discrete(labels = c('No', 'Yes')) +
labs(title = 'TB infection',
     subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                           analysis_set$TB), ]}))'),
     y = 'Number of sites') +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      plot.caption = element_text(size = 12),
      panel.grid = element_blank(),
      axis.title.x = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5)); tb

```



```
wilcox.test(Count ~ TB, data = analysis_set)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: Count by TB
## W = 13260, p-value = 0.2365
## alternative hypothesis: true location shift is not equal to 0
```

### 2.3.2 Continuous variables

```
bdi <- analysis_set %>%
  select(Count, BDI) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = BDI) +
  geom_point(size = 2,
            shape = 21,
            position = position_jitter(height = 0.2, width = 0.2),
            fill = '#FFFFFF',
            colour = '#000000',
            stroke = 0.8,
```

```

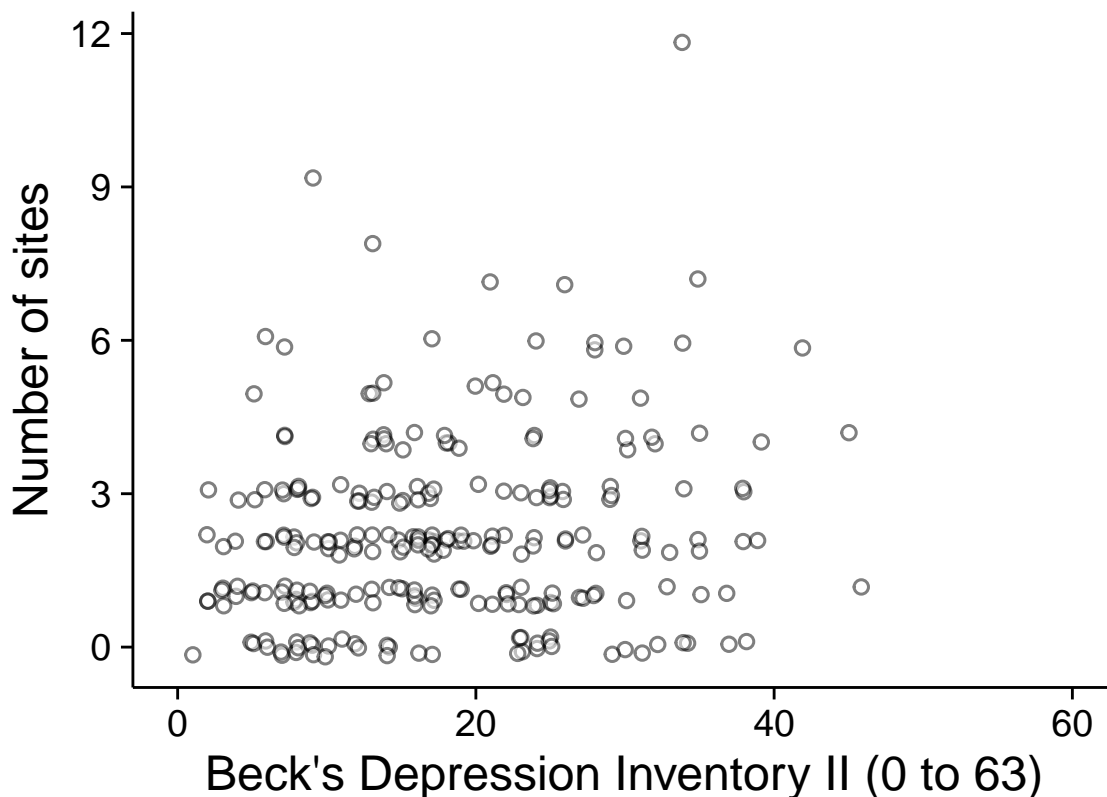
    alpha = 0.5) +
labs(title = 'Depression',
     subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
       analysis_set$BDI), ]})}')),
     x = "Beck's Depression Inventory II (0 to 63)",
     y = 'Number of sites') +
scale_x_continuous(limits = c(0, 60),
                  breaks = seq(0, 60, by = 20)) +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      plot.caption = element_text(size = 12),
      panel.grid = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5)); bdi

```

### 2.3.2.1 Beck's Depression Inventory II

## Depression

(Complete cases = 251)



```

cor.test(analysis_set$Count, analysis_set$BDI,
         method = 'spearman',
         exact = FALSE)

```

```
##
```

```
## Spearman's rank correlation rho
```

```
##
## data: analysis_set$Count and analysis_set$BDI
## S = 2241585, p-value = 0.01781
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.1494652
```

---

### 3 Session information

```
sessionInfo()

## R version 4.0.4 (2021-02-15)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] knitr_1.33      skimr_2.1.3     forcats_0.5.1  stringr_1.4.0
## [5] dplyr_1.0.7     purrr_0.3.4     readr_1.4.0    tidyr_1.1.3
## [9] tibble_3.1.3    ggplot2_3.3.5   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.1 xfun_0.25       repr_1.1.3      haven_2.3.1
## [5] colorspace_2.0-2 vctr_0.3.8      generics_0.1.0  htmltools_0.5.1.1
## [9] base64enc_0.1-3  yaml_2.2.1      utf8_1.2.2      rlang_0.4.11
## [13] pillar_1.6.2     glue_1.4.2      withr_2.4.2     DBI_1.1.1
## [17] dbplyr_2.1.0     modelr_0.1.8    readxl_1.3.1    lifecycle_1.0.0
## [21] munsell_0.5.0    gtable_0.3.0    cellranger_1.1.0 rvest_1.0.0
## [25] evaluate_0.14    labeling_0.4.2  fansi_0.5.0     highr_0.9
## [29] broom_0.7.9      Rcpp_1.0.7      scales_1.1.1    backports_1.2.1
## [33] jsonlite_1.7.2   farver_2.1.0    fs_1.5.0         hms_1.0.0
## [37] digest_0.6.27    stringi_1.7.3   grid_4.0.4      cli_3.0.1
## [41] tools_4.0.4      magrittr_2.0.1  crayon_1.4.1    pkgconfig_2.0.3
## [45] ellipsis_0.3.2   xml2_1.3.2      reprex_1.0.0    lubridate_1.7.10
## [49] assertthat_0.2.1 rmarkdown_2.10  httr_1.4.2      rstudioapi_0.13
## [53] R6_2.5.0         compiler_4.0.4
```