Supplement 3
Number of pain sites

Peter Kamerman

Last knitted: 16 October 2021

Contents

1 Import and check data

```
# Import
data <- read_rds('data-cleaned/data-pain-sites.rds')
demo <- read_rds('data-cleaned/data-demographics.rds')

# Check
```

```r
## Pain sites
dim(data)
```

```
## [1] 599  21
```

```r
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"
```

```r
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"~
```

```r
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
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  11
```

```
names(demo)
```

```
##  [1] "ID"                "Site"              "Sex"
##  [4] "Age"               "Employment_status" "CD4_recent"
##  [7] "ART_currently"     "Diabetes"          "TB"
## [10] "Education"         "BDI"
```

```
glimpse(demo)
```

```
## Rows: 599
## Columns: 11
## $ ID                <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB78"~
## $ Site              <chr> "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP"~
## $ Sex               <chr> "Female", "Female", "Female", "Female", "Female", "F~
## $ Age               <dbl> 36, 27, 39, 36, 31, 32, 28, 37, 31, 25, 31, 24, 35, ~
## $ Employment_status <chr> "Other", "Unemployed", "Other", "Unemployed", "Unemp~
## $ CD4_recent        <dbl> 391, 571, 591, 207, 126, 225, 543, 410, 74, 212, 579~
## $ ART_currently     <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Ye~
## $ Diabetes          <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ TB                <chr> "No", "No", "No", "No", "Yes", "No", "Yes", "No", "N~
## $ Education         <chr> "Tertiary", "Secondary", "Secondary", "Primary", "Se~
## $ BDI               <dbl> 23, 25, 5, 25, 24, 24, 8, 22, 9, 8, 8, 7, 3, 9, 26, ~
```

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

Table 3: Data summary

Name	Piped data
Number of rows	599
Number of columns	9
Column type frequency:	
factor	6
numeric	3
Group variables	None

Variable type: factor

skim_variable	missing	complete	n_unique	top_counts
Sex	0	599	2	Fem: 484, Mal: 115
Employment_status	49	550	4	Une: 333, Ful: 131, Par: 52, Oth: 34
ART_currently	5	594	2	Yes: 462, No: 132
Diabetes	241	358	2	No: 342, Yes: 16
TB	228	371	2	No: 261, Yes: 110
Education	37	562	3	Sec: 398, Pri: 99, Ter: 65

Variable type: numeric

skim_variable	missing	complete	mean	sd	p0	p25	p50	p75	p100
Age	8	591	37.23	9.06	19	31	36	42.00	76
CD4_recent	99	500	320.81	238.92	1	155	261	433.25	1232
BDI	348	251	17.91	9.74	1	10	17	25.00	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, Sex) %>%
  group_by(Sex) %>%
  skim()
```

2.2.2.1 By sex

Table 8: Data summary

Name	Piped data
Number of rows	599

Number of columns		2	

Column type frequency:	
numeric	1

Group variables	Sex

Variable type: numeric

skim_variable	Sex	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	Female	0	484	2.29	1.71	0	1	2	3	12
Count	Male	0	115	2.13	1.63	0	1	2	3	9

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

2.2.2.2 By ART

Table 10: Data summary

Name	Piped data
Number of rows	594
Number of columns	2

Column type frequency:	
numeric	1

Group variables	ART_currently

Variable type: numeric

skim_variable	ART_currently	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	No	0	132	2.21	1.65	0	1	2	3	12
Count	Yes	0	462	2.28	1.71	0	1	2	3	9

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

2.2.2.3 By education

Table 12: Data summary

Name	Piped data
Number of rows	562
Number of columns	2
Column type frequency:	
numeric	1
Group variables	Education

Variable type: numeric

skim_variable	Education	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	Primary	0	99	2.38	1.64	0	1	2	3	8
Count	Secondary	0	398	2.26	1.75	0	1	2	3	12
Count	Tertiary	0	65	2.06	1.52	0	1	2	3	6

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

2.2.2.4 By employment status

Table 14: Data summary

Name	Piped data
Number of rows	550
Number of columns	2
Column type frequency:	
numeric	1
Group variables	Employment_status

Variable type: numeric

skim_variable	Employment_status	missing	complete	mean	sd	p0	p25	p50	p75	p100
Count	Full-time work	0	131	2.21	1.46	0	1	2	3	6
Count	Other	0	34	2.18	1.64	0	1	2	4	5
Count	Part-time work	0	52	2.50	1.89	0	1	2	3	9
Count	Unemployed	0	333	2.23	1.77	0	1	2	3	12

2.3 Exploratory plots

2.3.1 Overall count frequency

```r
all <- analysis_set %>%
    select(Count) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(Count) +
    geom_histogram(binwidth = 1,
                   colour = '#FFFFFF') +
    labs(title = 'All participants',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Count), ])})'),
         x = 'Number of sites',
         y = 'Count') +
    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)); all
```


8
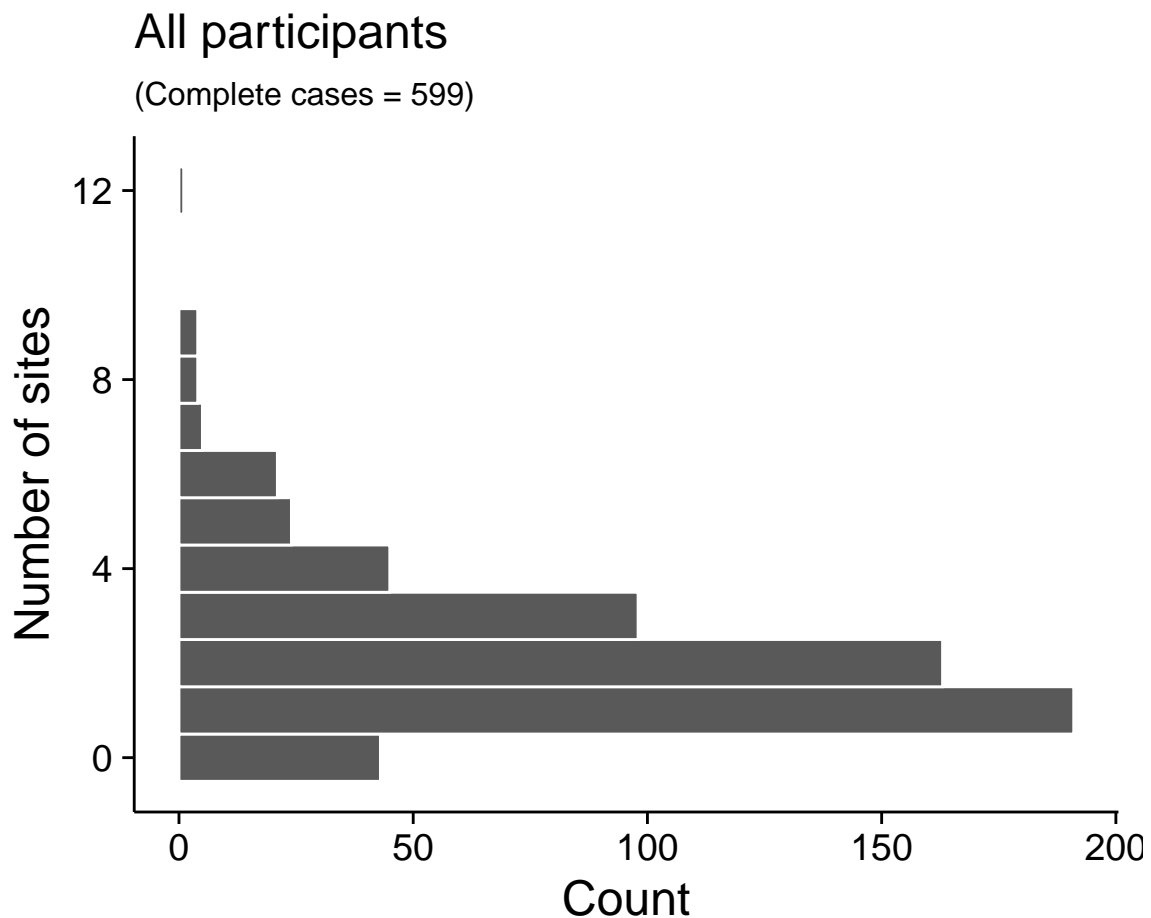
```r
all <- analysis_set %>%
    select(Count) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = 'All participants') +
    geom_point(size = 2,
               shape = 21,
               position = position_jitter(height = 0.2, width = 0.2),
               fill = '#FFFFFF',
               colour = '#000000',
               stroke = 0.8,
               alpha = 0.5) +
    geom_boxplot(outlier.colour = '#FFFFFF',
                 outlier.size = 0,
                 width = 0.5,
                 size = 0.8,
                 alpha = 0) +
    labs(title = 'All participants',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Count), ])})'),
         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)); all
```

All participants
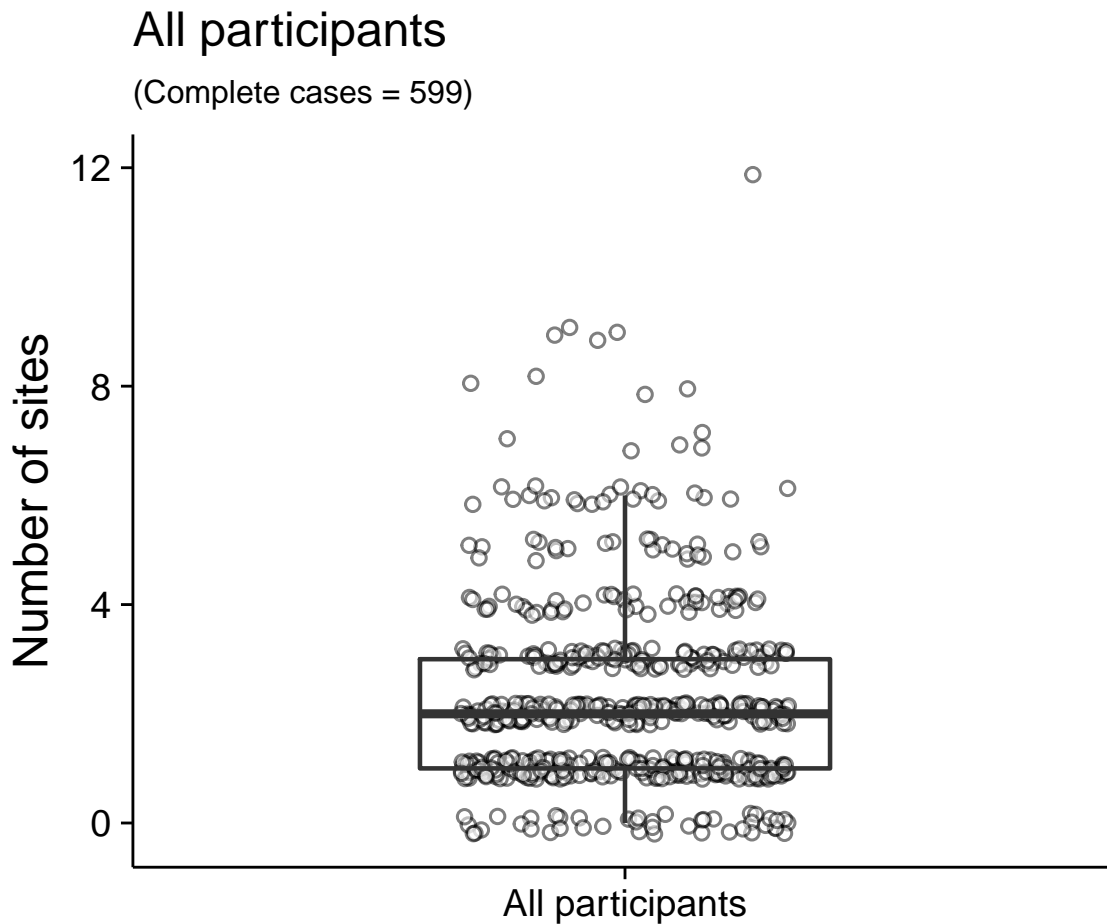
(Complete cases = 599)

2.3.2 For each categorical variable
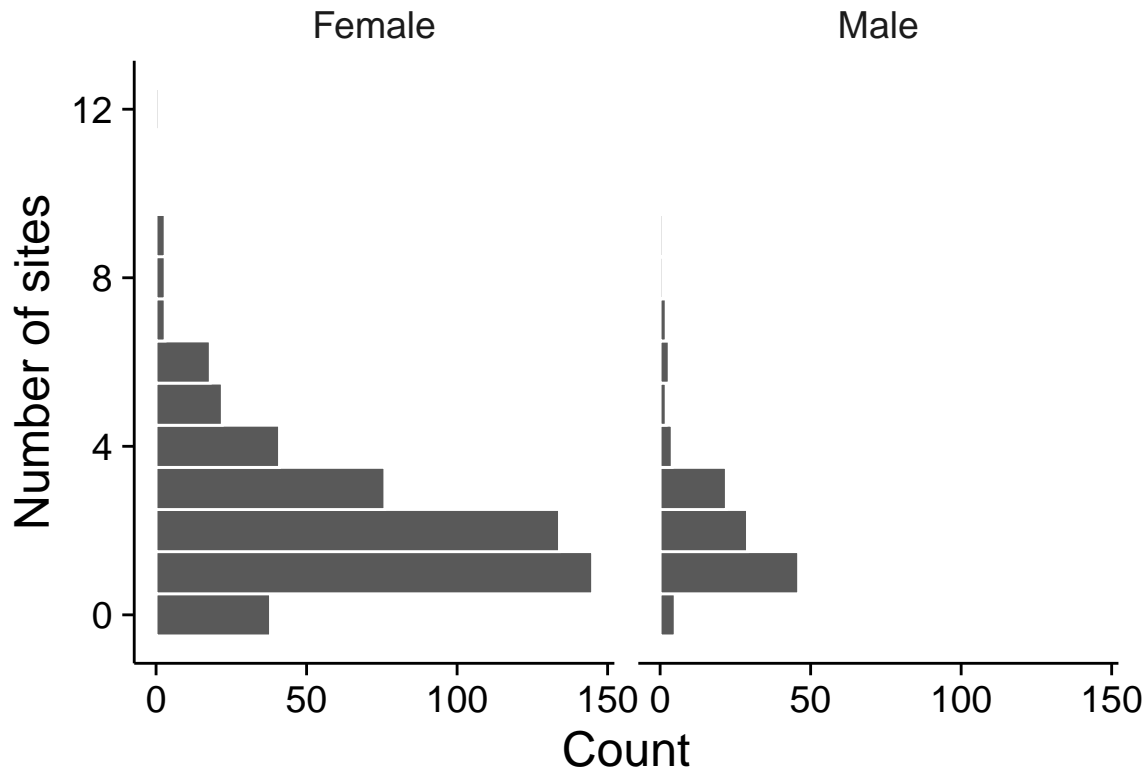
```
analysis_set %>%
    select(Count, Sex) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(Count) +
    geom_histogram(binwidth = 1,
                   colour = '#FFFFFF') +
    labs(title = 'Sex',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Sex), ])})'),
         x = 'Number of sites',
         y = 'Count') +
    facet_wrap(~Sex) +
    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))
```

2.3.2.1   Count by sex

Sex

(Complete cases = 599)


```
sex <- analysis_set %>%
    select(Count, Sex) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = Sex) +
    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('Female', 'Male')) +
    labs(title = 'Sex',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Sex), ])})'),
```
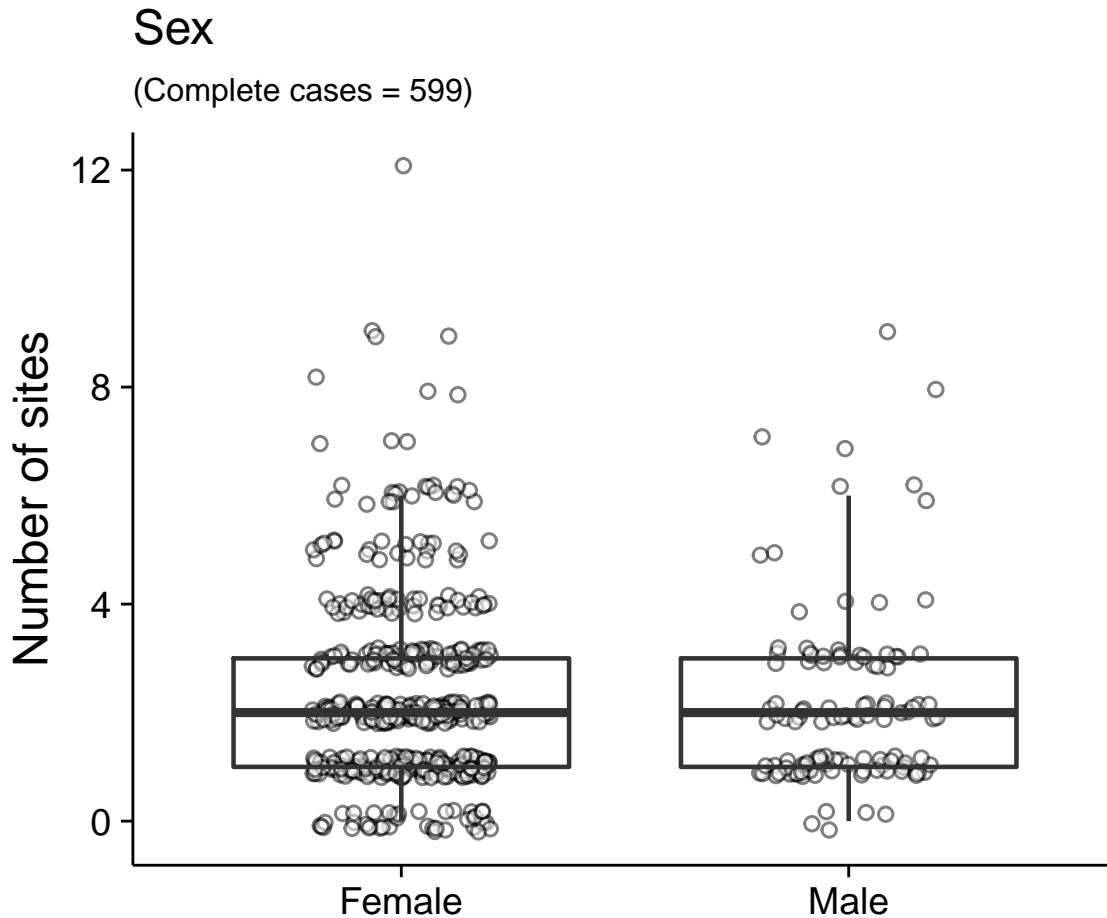
```
        y = 'Number of sites') +
    theme_minimal(base_size = 18) +
    theme(plot.title = element_text(size = 18),
          plot.subtitle = element_text(size = 12),
          plot.caption = element_text(size = 12),
          panel.grid = element_blank(),
          axis.title.x = element_blank(),
          axis.text = element_text(colour = '#000000'),
          axis.line = element_line(size = 0.5),
          axis.ticks = element_line(size = 0.5)); sex
```

Sex

(Complete cases = 599)


```
analysis_set %>%
    select(Count, ART_currently) %>%
    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 = 'Currently on ART',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
```
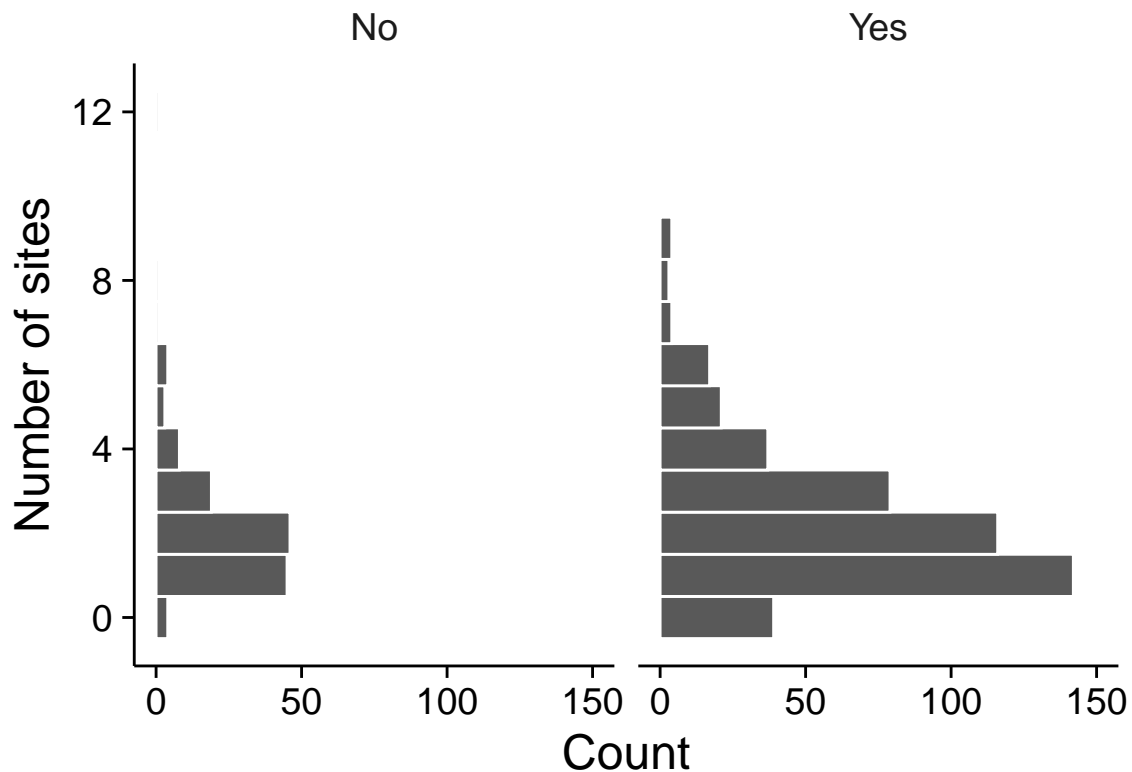
```
                             analysis_set$ART_currently), ])})'),
        x = 'Number of sites',
        y = 'Count') +
  facet_wrap(~ART_currently) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))
```

2.3.2.2 Count by ART

Currently on ART

(Complete cases = 594)


```
haart <- analysis_set %>%
    select(Count, ART_currently) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = ART_currently) +
    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 = 'Currently on ART',
     subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                          analysis_set$ART_currently), ])})'),
     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)); haart
```
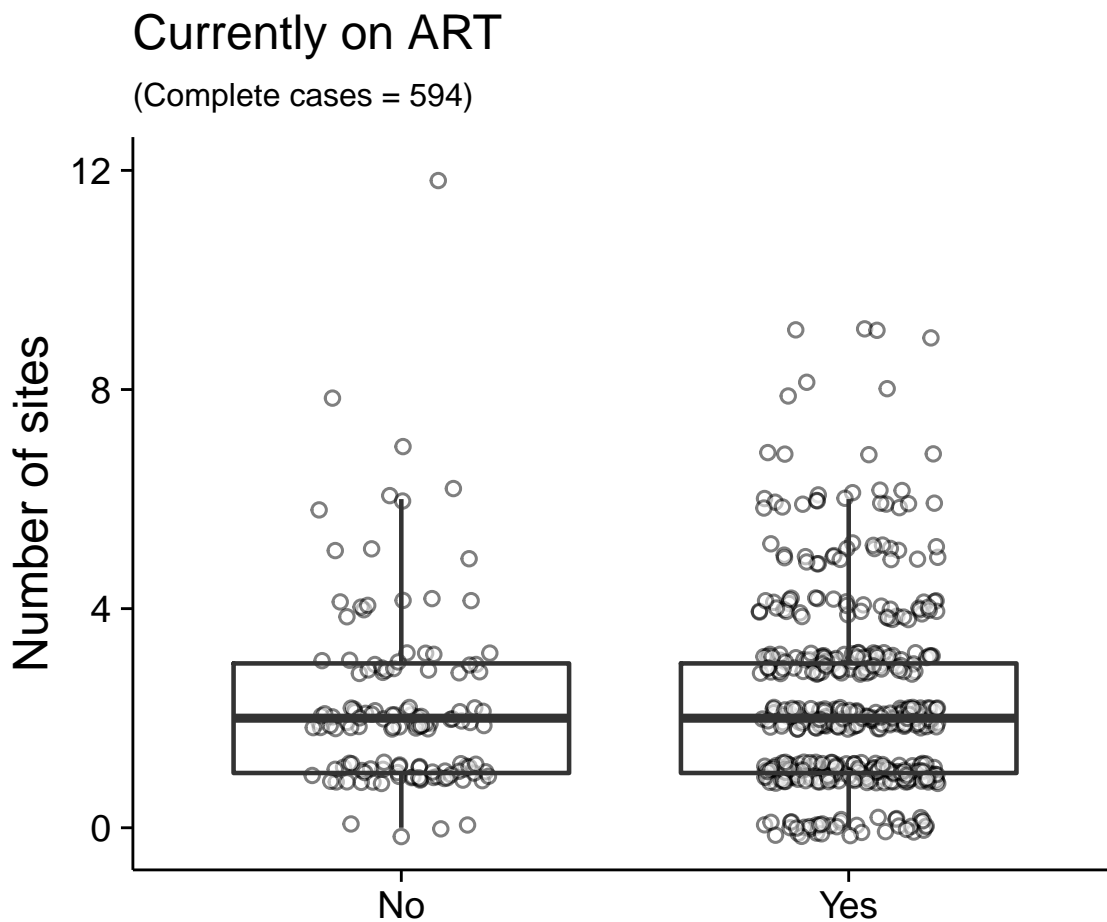

Currently on ART

(Complete cases = 594)
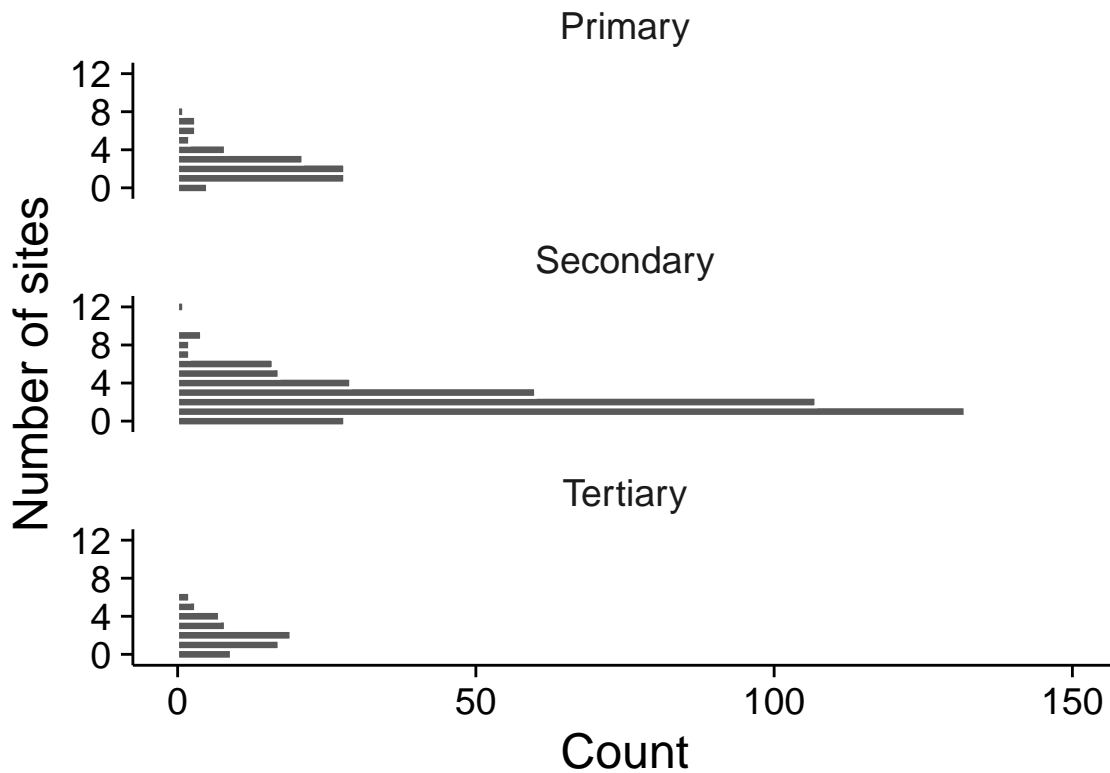
```r
analysis_set %>%
    select(Count, Education) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(Count) +
    geom_histogram(binwidth = 1,
                   colour = '#FFFFFF') +
    scale_y_continuous(limits = c(0, 150),
                       breaks = c(0, 50, 100, 150)) +
    labs(title = 'Level of education',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Education), ])})'),
         x = 'Number of sites',
         y = 'Count') +
    facet_wrap(~Education, ncol = 1) +
    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))
```

2.3.2.3 Count education

Level of education

(Complete cases = 562)


```r
edu <- analysis_set %>%
    select(Count, Education) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = Education) +
    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('Primary', 'Secondary',
                                'Tertiary')) +
    labs(title = 'Level of Education',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Education), ])})'),
         y = 'Number of sites') +
    theme_minimal(base_size = 18) +
    theme(plot.title = element_text(size = 18,
```
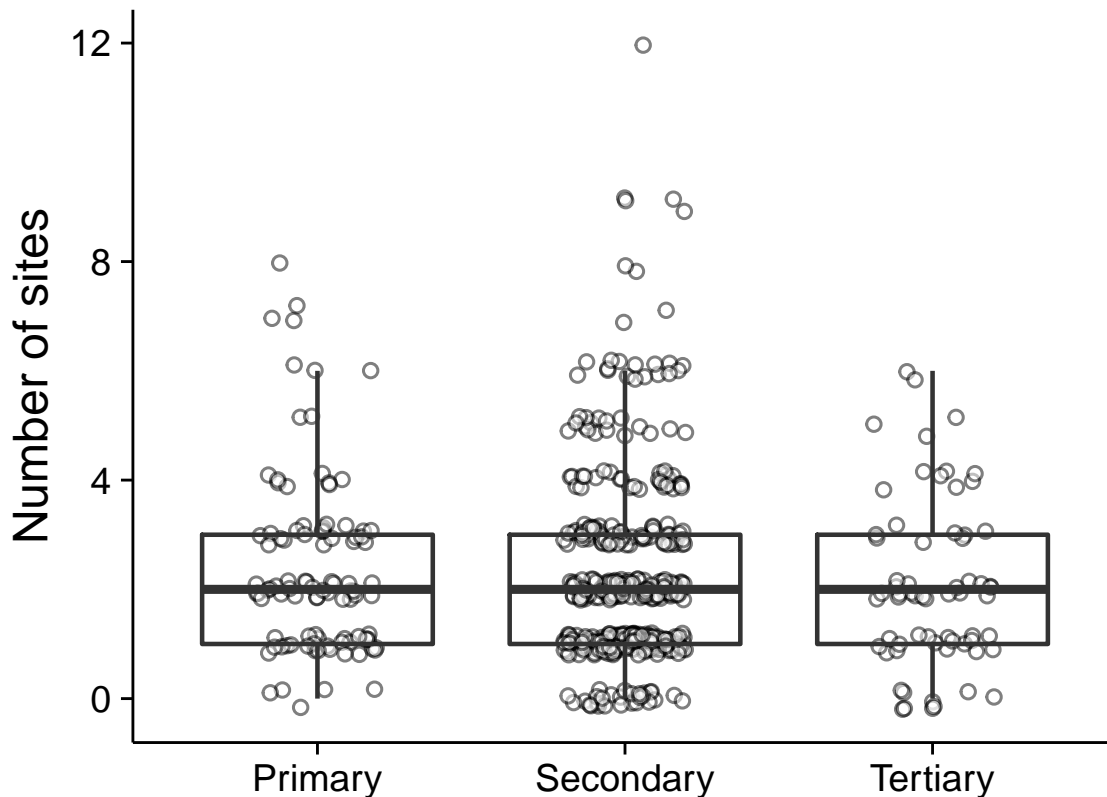
```
    plot.subtitle = element_text(size = 12),
    plot.caption = element_text(size = 12),
    panel.grid = element_blank(),
    axis.title.x = element_blank(),
    axis.text = element_text(colour = '#000000'),
    axis.line = element_line(size = 0.5),
    axis.ticks = element_line(size = 0.5)); edu
```

Level of Education

(Complete cases = 562)


```
analysis_set %>%
    select(Count, Employment_status) %>%
    filter(complete.cases(.)) %>%
    mutate(Employment_status = factor(Employment_status,
                                      levels = c('Full-time work',
                                                 'Part-time work',
                                                 'Other',
                                                 'Unemployed'),
                                      ordered = TRUE)) %>%
    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 = 'Employment status',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                              analysis_set$Employment_status), ])})'),
         caption = 'Other: were receiving social grants or were students',
         x = 'Number of sites',
         y = 'Count') +
    facet_wrap(~Employment_status, ncol = 2) +
    coord_flip() +
    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))
```
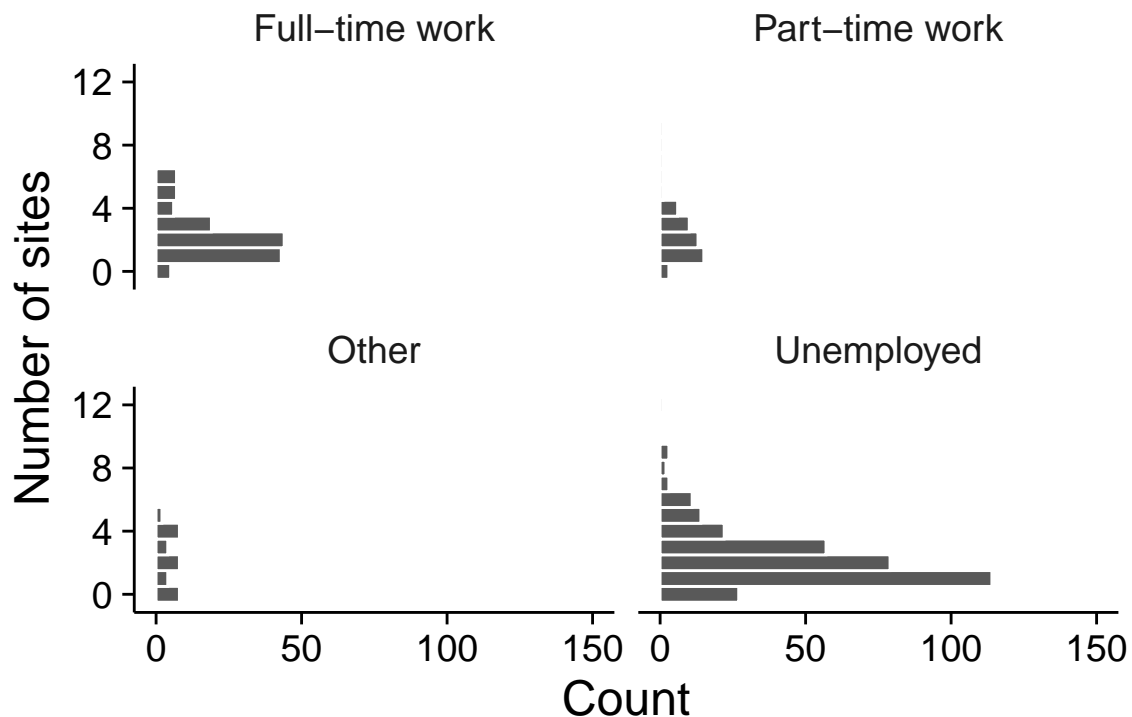
2.3.2.4 Count by employment status

Employment status

(Complete cases = 550)

Other: were receiving social grants or were students

```
employment <- analysis_set %>%
    select(Count, Employment_status) %>%
    filter(complete.cases(.)) %>%
    mutate(Employment_status = factor(Employment_status,
```

```
                                       levels = c('Full-time work',
                                                  'Part-time work',
                                                  'Other',
                                                  'Unemployed'),
                                       ordered = TRUE)) %>%
ggplot(data = .) +
aes(y = Count,
    x = Employment_status) +
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('Full-time\nwork', 'Part-time\nwork',
                            'Other*', 'Unemployed')) +
labs(title = 'Employment status',
     subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                         analysis_set$Employment_status), ])})'),
     caption = '*Other: were receiving social grants or were students',
     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)); employment
```
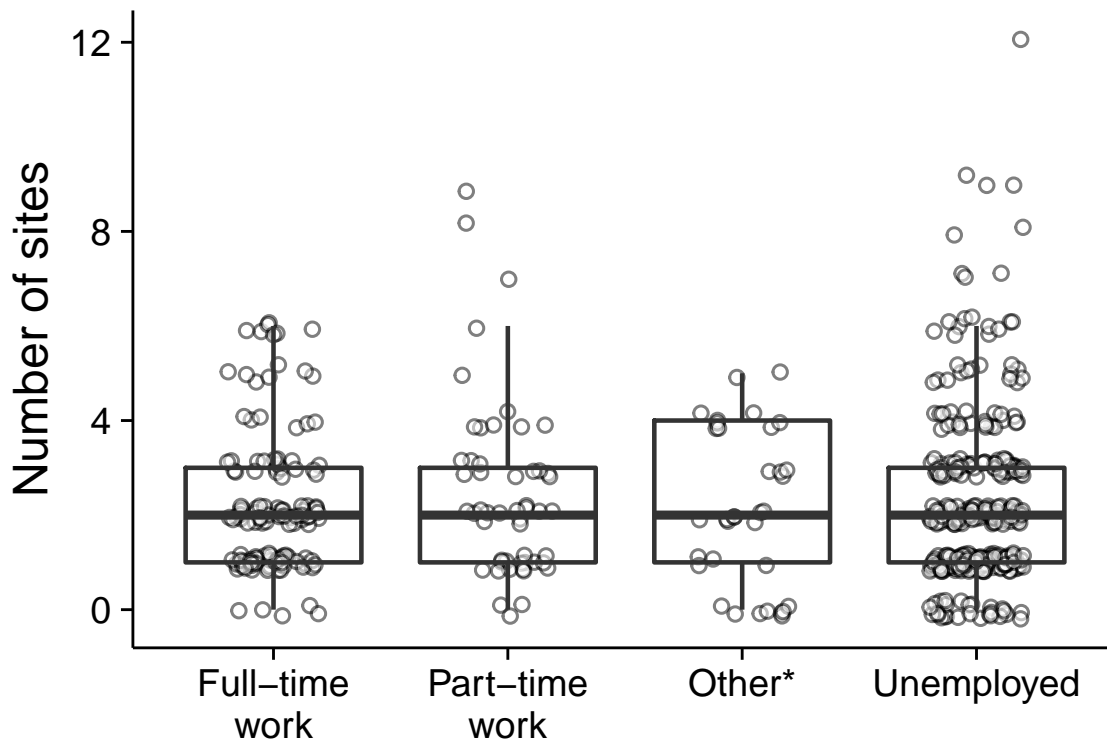
Employment status

(Complete cases = 550)

*Other: were receiving social grants or were students

2.3.3 Continuous variables

```r
age <- analysis_set %>%
    select(Count, Age) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = Age) +
    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 = 'Age',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$Age), ])})'),
         x = 'Age (years)',
         y = 'Number of sites') +
    scale_x_continuous(limits = c(18, 80),
                       breaks = seq(20, 80, 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)); age
```

2.3.3.1 Age

Age
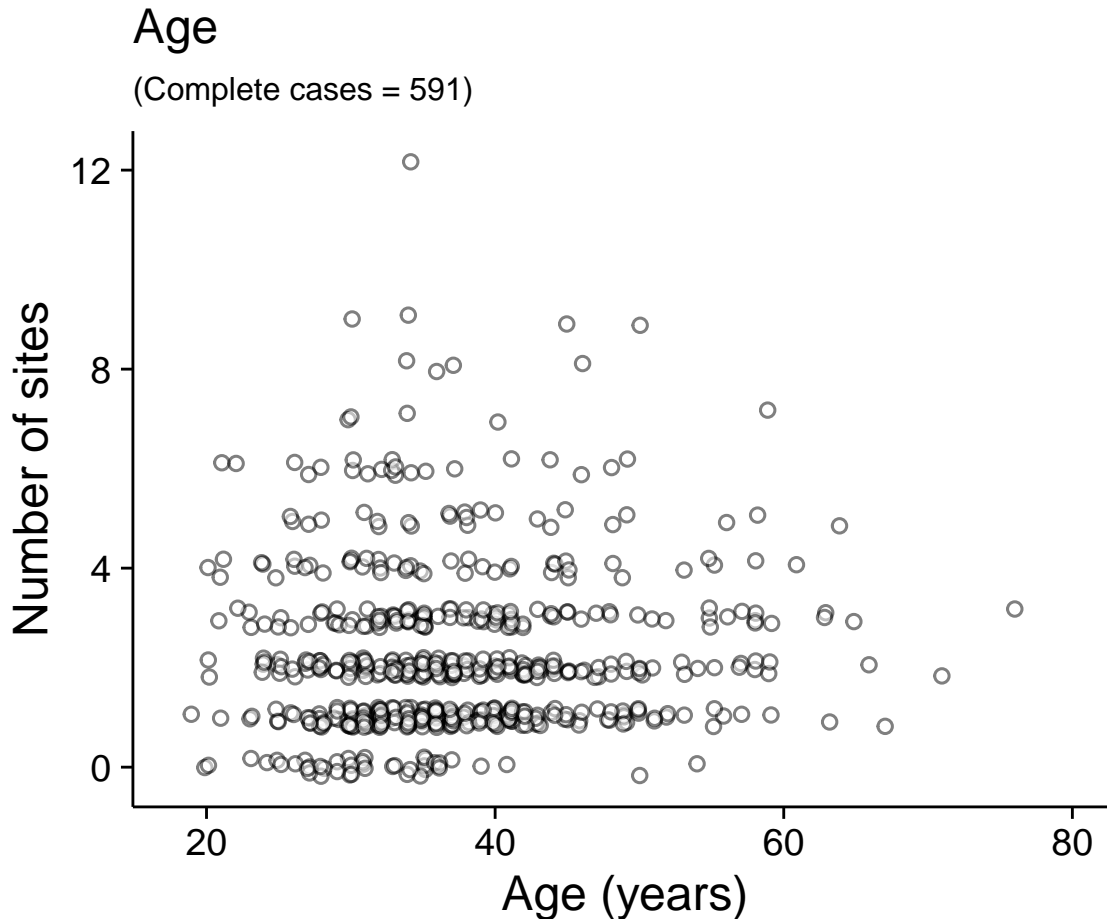
(Complete cases = 591)


```
cd4 <- analysis_set %>%
    select(Count, CD4_recent) %>%
    filter(complete.cases(.)) %>%
    ggplot(data = .) +
    aes(y = Count,
        x = CD4_recent) +
    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 = 'Recent CD4 T-cell count',
         subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                             analysis_set$CD4_recent), ])})'),
         x = expression('CD4 T-cell count (cells.mm'^-3*')'),
         y = 'Number of sites') +
    scale_x_continuous(limits = c(0, 1400),
                       breaks = seq(0, 1400, by = 400)) +
    scale_y_continuous(breaks = seq(0, 12, 4)) +
    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)); cd4
```
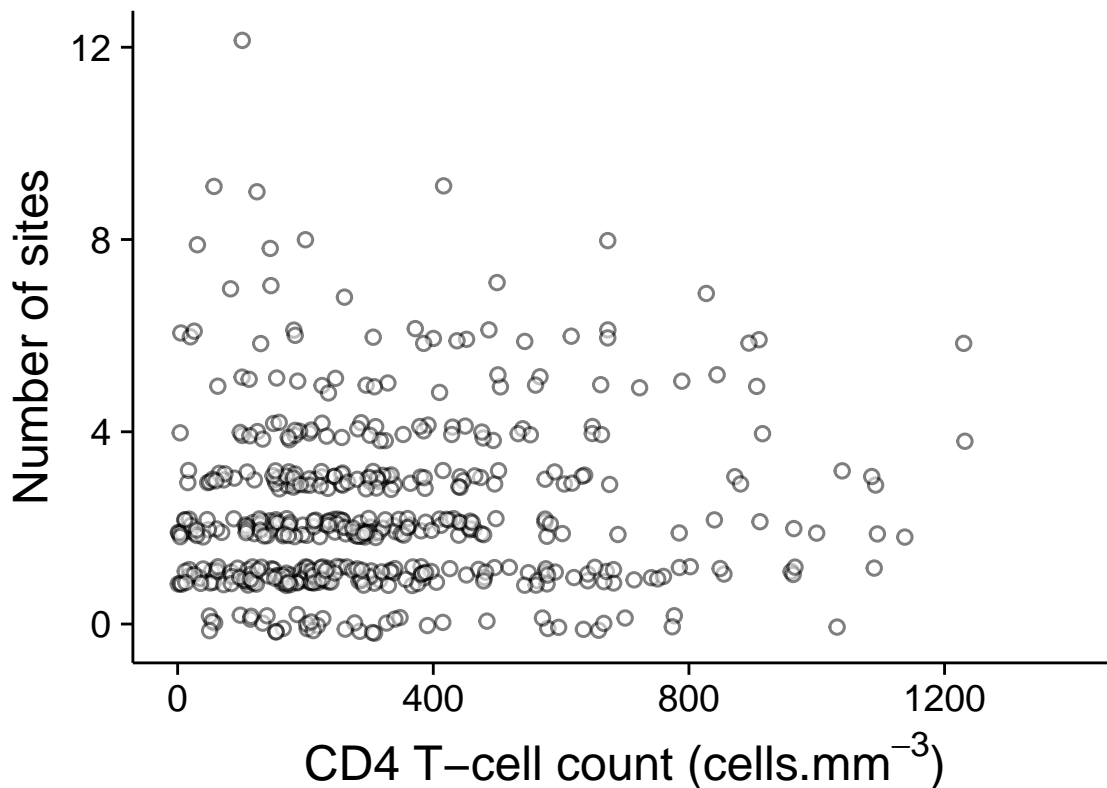
2.3.3.2   Current CD4 T-cell count

Recent CD4 T–cell count

(Complete cases = 500)

2.4 Regression analysis

2.4.1 Process data

```
analysis_set <- analysis_set %>%
    # Re-factor Employment_status
    mutate(Employment_status = case_when(
        Employment_status == 'Unemployed' ~ ' 1_unemployed',
        Employment_status == 'Part-time work' ~ ' 3_employed PT',
        Employment_status == 'Full-time work' ~ ' 2_employed FT',
        Employment_status == 'Other' ~ ' 4_other'
    )) %>%
    filter(complete.cases(.))
```

2.4.2 Negative binomial regression for count data

```
# Full model
mod.nb <- glm.nb(Count ~ Sex +
                    Age +
                    CD4_recent +
                    Employment_status +
                    Education +
                    ART_currently,
                 data = analysis_set)

# Summary of coefficients
cbind(Estimate = coef(mod.nb)[-1],
      confint(mod.nb)[-1, ],
      summary(mod.nb)$coefficients[-1, 3:4]) %>%
    kable(caption = 'Coefficients and 95% CI',
          digits = 3,
          col.names = c('Estimate', 'Lower 95%CI', 'Upper 95%CI',
                        'z-value', 'P-value'))
```

Table 16: Coefficients and 95% CI

	Estimate	Lower 95%CI	Upper 95%CI	z-value	P-value
SexMale	-0.259	-0.816	0.269	-0.945	0.345
Age	0.011	-0.006	0.027	1.272	0.203
CD4_recent	0.000	-0.001	0.000	-0.627	0.530
Employment_status 2_employed FT	-0.139	-0.519	0.234	-0.722	0.470
Employment_status 3_employed PT	0.119	-0.268	0.499	0.608	0.543
Employment_status 4_other	-0.040	-0.380	0.293	-0.234	0.815
EducationSecondary	-0.149	-0.495	0.202	-0.839	0.401
EducationTertiary	-0.413	-0.867	0.039	-1.784	0.074
ART_currentlyYes	-0.276	-0.653	0.105	-1.431	0.153

```
# Summary of exponentiated coefficients (incidence rate ratios)
cbind(Estimate = exp(coef(mod.nb))[-1],
      exp(confint(mod.nb))[-1, ],
      summary(mod.nb)$coefficients[-1, 3:4]) %>%
```

```
kable(caption = 'Exponentiate coefficients and 95% CI (incidence rate ratios)',
      digits = 3,
      col.names = c('Estimate', 'Lower 95%CI', 'Upper 95%CI',
                    'z-value', 'P-value'))
```

Table 17: Exponentiate coefficients and 95% CI (incidence rate ratios)

	Estimate	Lower 95%CI	Upper 95%CI	z-value	P-value
SexMale	0.771	0.442	1.309	-0.945	0.345
Age	1.011	0.994	1.028	1.272	0.203
CD4_recent	1.000	0.999	1.000	-0.627	0.530
Employment_status 2_employed FT	0.871	0.595	1.263	-0.722	0.470
Employment_status 3_employed PT	1.126	0.765	1.648	0.608	0.543
Employment_status 4_other	0.961	0.684	1.341	-0.234	0.815
EducationSecondary	0.862	0.609	1.224	-0.839	0.401
EducationTertiary	0.662	0.420	1.040	-1.784	0.074
ART_currentlyYes	0.759	0.520	1.110	-1.431	0.153

2.5 Publication plot

```
# Fix figures for patchwork plot
all2 <- all

sex2 <- sex +
  theme(axis.title.y = element_blank())

haart2 <- haart

employment2 <- employment

edu2 <- edu +
  theme(axis.title.y = element_blank())

age2 <- age +
  theme(axis.title.y = element_blank())

cd42 <- cd4

pubs <- all2 + sex2 + haart2 + edu2 + employment2 + age2 + cd42 +
  plot_layout(ncol = 2)

ggsave(filename = 'figures/figure_2.png',
       width = 10,
       height = 18)
```

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] patchwork_1.1.1 knitr_1.33      skimr_2.1.3     forcats_0.5.1
##  [5] stringr_1.4.0   dplyr_1.0.7     purrr_0.3.4     readr_1.4.0
##  [9] tidyr_1.1.3     tibble_3.1.3    ggplot2_3.3.5   tidyverse_1.3.0
## [13] MASS_7.3-53.1
##
## 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 vctrs_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
```