Owen et al. - Figure 1

This script will reproduce Figure 1 from *Owen et al. - Sample size planning for insect critical thermal limits studies.*

Load required packages

```r
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse,
               tidyr,
               readr,
               here,
               cowplot,
               janitor)
```

Set ggplot theme (makes nice plots)

```r
theme_set(theme_classic() +
            theme(panel.border = element_rect(colour = "black", fill = NA),
                  axis.text = element_text(colour = "black"),
                  axis.title.x = element_text(margin = unit(c(2, 0, 0, 0), "mm")),
                  axis.title.y = element_text(margin = unit(c(0, 4, 0, 0), "mm")),
                  legend.position = "none"))
```

Load GlobalTherm database raw data

```r
# Also, load raw data (without CI's)
raw_data <- readr::read_csv2(here::here("./data_raw/GlobalTherm_upload_10_11_17.csv"))

# Clean df, and keep only insect records
raw_data <- raw_data %>%
  janitor::clean_names() %>%
  # Filter to only keep the studies on insects
  dplyr::filter(class == "Insecta")
```

Split data in CTmin and CTmax

```r
# Make a CTmax df
max_data <- raw_data %>%
  dplyr::select(genus,
```

```r
                 species,
                 n,
                 tmax,
                 max_metric,
                 error,
                 error_measure) %>%
  dplyr::filter(max_metric == "ctmax")

# Make a CTmin df
min_data <- raw_data %>%
  dplyr::select(genus,
                 species,
                 n,
                 tmin,
                 min_metric,
                 error,
                 error_measure) %>%
  dplyr::filter(min_metric == "ctmin")
```

CTMax analysis

```r
# Calculate number of studies
max_data %>%
  tidyr::drop_na(n) %>%
  nrow()
```

```
## [1] 76
```

```r
# Calculate range and median sample size used for CTmax studies
max_data %>%
  tidyr::drop_na(n) %>%
  dplyr::summarise(median = median(n),
                   min = min(n),
                   max = max(n),
                   sd = sd(n),
                   nn = n(),
                   se = sd/sqrt(nn))
```

```
## # A tibble: 1 x 6
##   median   min   max    sd    nn    se
##    <dbl> <dbl> <dbl> <dbl> <int> <dbl>
## 1     18     2  5550 1066.    76  122.
```

```r
# Calculate proportion of studies using more than n = 30
max_data %>%
  tidyr::drop_na(n) %>%
  dplyr::mutate(below_30 = dplyr::case_when(
    n > 30 ~ 1,
    TRUE ~ 0
  )) %>%
```

```
  dplyr::mutate(nn = n()) %>%
  dplyr::mutate(n_above_30 = sum(below_30),
                prop_above_30 = sum(below_30)/nn) %>%
  dplyr::slice(1) %>%
  dplyr::select(n_above_30, prop_above_30)
```

```
## # A tibble: 1 x 2
##   n_above_30 prop_above_30
##        <dbl>         <dbl>
## 1         13         0.171
```

```
# Calculate proportion of studies using more than n = 120
max_data %>%
  tidyr::drop_na(n) %>%
  dplyr::mutate(below_100 = dplyr::case_when(
    n > 100 ~ 1,
    TRUE ~ 0
  )) %>%
  dplyr::mutate(nn = n()) %>%
  dplyr::mutate(n_above_100 = sum(below_100),
                prop_above_100 = sum(below_100)/nn) %>%
  dplyr::slice(1) %>%
  dplyr::select(n_above_100, prop_above_100)
```

```
## # A tibble: 1 x 2
##   n_above_100 prop_above_100
##         <dbl>          <dbl>
## 1           7         0.0921
```

```
# Plot frequency distribution of CTmax sample sizes
plot_ctmax <- ggplot(data = max_data, aes(x = n)) +
  geom_histogram(fill = "gray60",
                 binwidth = 5,
                 center = 1) +
  scale_x_continuous(breaks = seq(0, 70, 10),
                     limits = c(0, 70),
                     expand = c(0,0)) +
  scale_y_continuous(breaks = seq(0, 30, 5),
                     limits = c(0, 30),
                     expand = c(0,0)) +
  labs(x = "Sample size",
       y = "Frequency",
       subtitle = "(b) CTmax") +
  annotate("text",
           x = 69,
           y = 27.5,
           label = "n = 76 studies",
           hjust = 1) +
  annotate("text",
           x = 69,
           y = 25,
           label = "n > 70 for 7 studies (9%)",
```

```
            hjust = 1)
plot_ctmax
```
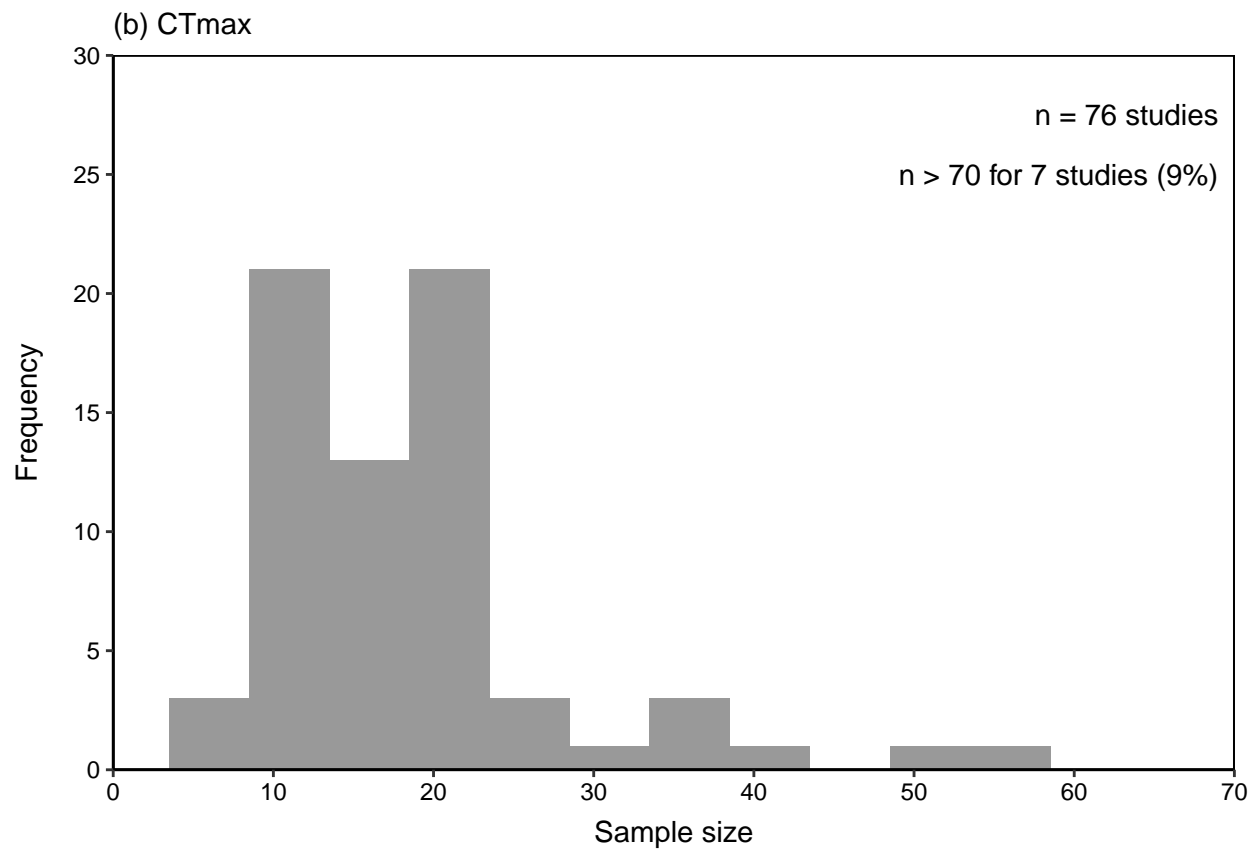
(b) CTmax

CTMin analysis

```
# Calculate number of studies
min_data %>%
  tidyr::drop_na(n) %>%
  nrow()
```

```
## [1] 33
```

```
# Calculate range and median sample size used for CTmax studies
min_data %>%
  tidyr::drop_na(n) %>%
  dplyr::summarise(median = median(n),
                   min = min(n),
                   max = max(n),
                   sd = sd(n),
                   nn = n(),
                   se = sd/sqrt(nn))
```

```
## # A tibble: 1 x 6
```

4

```
##    median   min   max    sd    nn    se
##      <dbl> <dbl> <dbl> <dbl> <int> <dbl>
## 1      15     8   500  84.9    33  14.8
```

```r
# Calculate proportion of studies using more than n = 30
min_data %>%
  tidyr::drop_na(n) %>%
  dplyr::mutate(below_30 = dplyr::case_when(
    n > 30 ~ 1,
    TRUE ~ 0
  )) %>%
  dplyr::mutate(nn = n()) %>%
  dplyr::mutate(n_above_30 = sum(below_30),
                prop_above_30 = sum(below_30)/nn) %>%
  dplyr::slice(1) %>%
  dplyr::select(n_above_30, prop_above_30)
```

```
## # A tibble: 1 x 2
##   n_above_30 prop_above_30
##        <dbl>         <dbl>
## 1          2        0.0606
```

```r
# Calculate proportion of studies using more than n = 120
min_data %>%
  tidyr::drop_na(n) %>%
  dplyr::mutate(below_100 = dplyr::case_when(
    n > 100 ~ 1,
    TRUE ~ 0
  )) %>%
  dplyr::mutate(nn = n()) %>%
  dplyr::mutate(n_above_100 = sum(below_100),
                prop_above_100 = sum(below_100)/nn) %>%
  dplyr::slice(1) %>%
  dplyr::select(n_above_100, prop_above_100)
```

```
## # A tibble: 1 x 2
##   n_above_100 prop_above_100
##         <dbl>          <dbl>
## 1           1         0.0303
```

```r
# Plot frequency distribution of CTmin sample sizes
plot_ctmin <- ggplot(data = min_data, aes(x = n)) +
  geom_histogram(fill = "gray60",
                 binwidth = 5,
                 center = 1) +
  scale_x_continuous(breaks = seq(0, 70, 10),
                     limits = c(0, 70),
                     expand = c(0,0)) +
  scale_y_continuous(breaks = seq(0, 30, 5),
                     limits = c(0, 30),
                     expand = c(0,0)) +
  labs(x = "Sample size",
```

```
      y = "Frequency",
      subtitle = "(a) CTmin") +
 annotate("text",
          x = 69,
          y = 27.5,
          label = "n = 33 studies",
          hjust = 1) +
 annotate("text",
          x = 69,
          y = 25,
          label = "n > 70 for 1 study (3%)",
          hjust = 1)
plot_ctmin
```
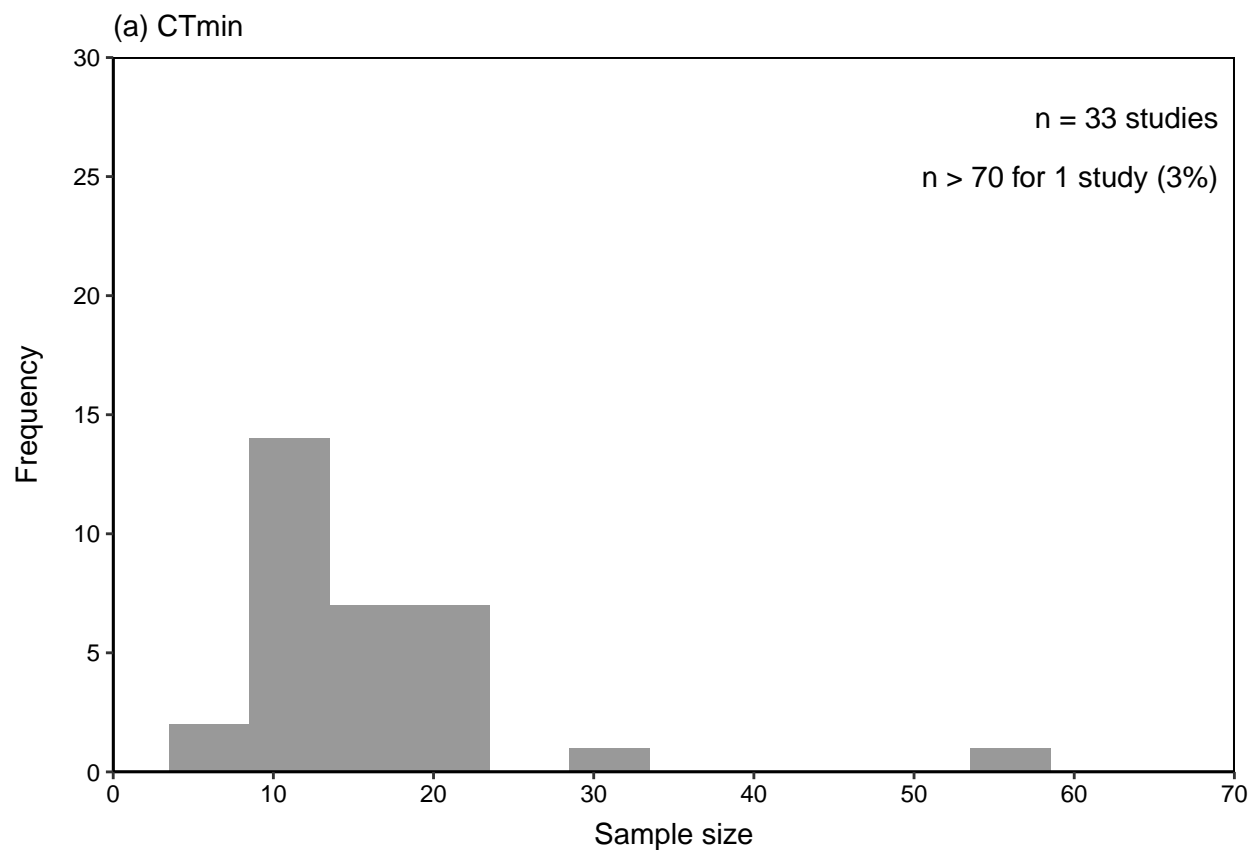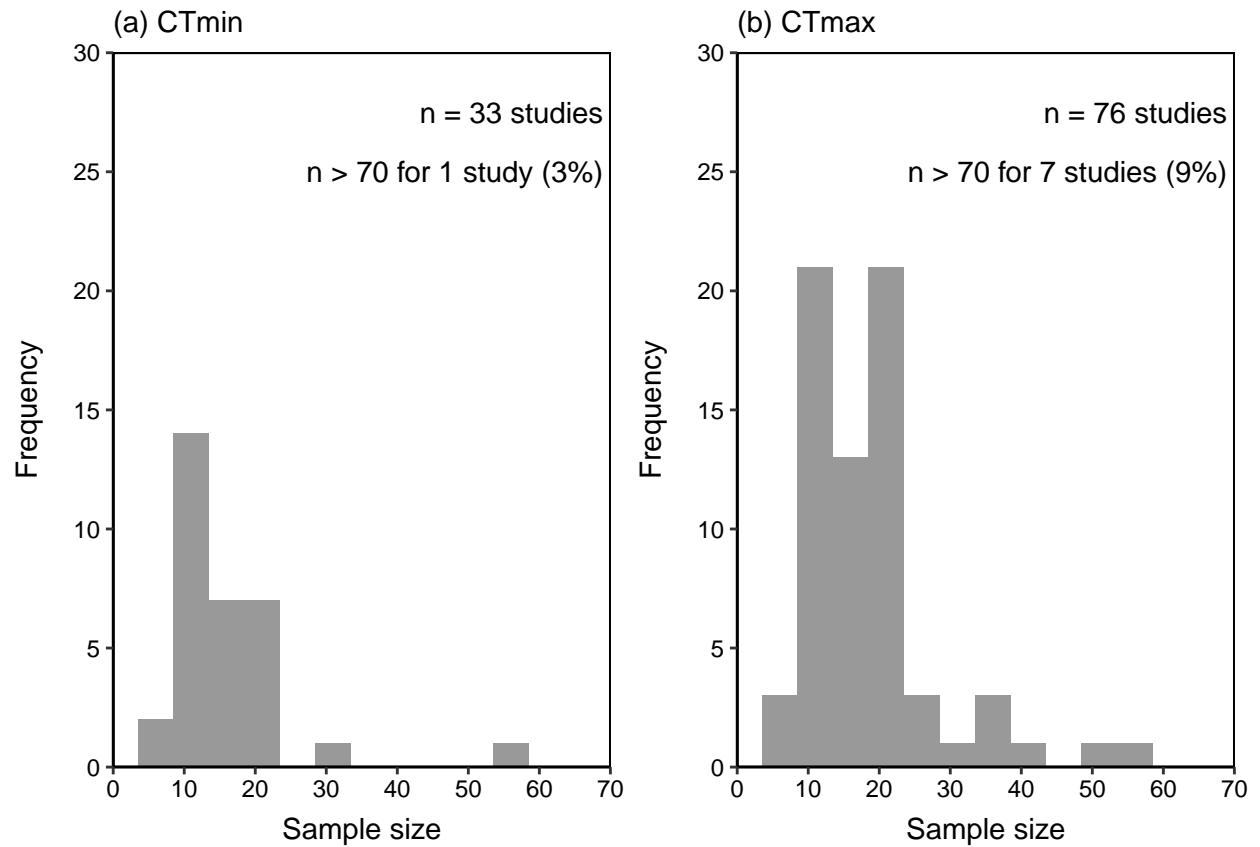

(a) CTmin

Put the two plots together in one graph

```
# Put together
fig1 <- cowplot::plot_grid(plot_ctmin,
                  plot_ctmax,
                  ncol = 2)
fig1
```

Save plot to PC

```r
# Save figure to file
ggplot2::ggsave(here::here("./figures/figure_1.png"),
                width = 8,
                height = 5)
```