

# Explore landscape covariates

Aidan Brushett

2025-05-02

## Contents

<b>Before you begin</b>	<b>1</b>
<b>0. Setup</b>	<b>2</b>
<b>1. Import data</b>	<b>2</b>
<b>2. Data distribution</b>	<b>3</b>
2.1. Histograms at 500 meters: . . . . .	3
2.2. Box and whiskers at 500 meters: . . . . .	37
2.3. Histograms at 3500 meters: . . . . .	71
2.4. Box and whiskers at 3500 meters: . . . . .	105
<b>3. Correlation</b>	<b>139</b>
<b>4. Summary table of covariates</b>	<b>140</b>

## Before you begin

This script is number 3 of 6 in a series of scripts used to replicate the analyses presented in the paper: “Life on the edge: Industrial footprint and edge effects variably affect the distribution of a boreal small mammal”

This script was used to conduct the exploratory analysis for all potential landscape covariates in our dataset prior to modelling. **The raw spatial data used to develop landcover metrics is not available on GitHub but can be shared by the authors upon request.**

When running these scripts, please ensure that you have downloaded the complete GitHub repository. This will ensure you have all the files, data, and proper folder structure you will need to run this code and associated analyses.

Also make sure you open RStudio through the R project (OSM\_red\_squirrel\_distribution.Rproj). This will automatically set your working directory to the correct place (wherever you saved the repository) and ensure you don't have to change the file paths for some of the data. This analysis was initially run in R v4.3.0. If you have any questions or concerns, please contact one of the authors:

Aidan Brushett M.Sc. Student University of Victoria  
School of Environmental Studies  
Email: aidanbrushett@uvic.ca

Emerald Arthurs M.Sc. Student University of Victoria  
School of Environmental Studies

---

## 0. Setup

```
rm(list=ls()) # start fresh :)

library(tidyverse)
library(ggplot2)
library(writexl)

# These are functions that can sometimes get masked by other packages. Shouldn't be an issue and
# ↪ could be removed.
select <- dplyr::select
filter <- dplyr::filter
mutate <- dplyr::mutate
summarize <- dplyr::summarize
# Cheeky function that means "not in"
`%nin%` = Negate(`%in%`)
```

## 1. Import data

```
covs <- read_csv("../data/processed/OSM_all_covariates_HFI_SBFI_final.csv")

## Rows: 9460 Columns: 55
## -- Column specification -----
## Delimiter: ","
## chr (2): array, site
## dbl (53): array_year, lat, long, easting_12n, northing_12n, buffer_dist, cfi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
names(covs)
```

```
## [1] "array"           "site"            "array_year"
## [4] "lat"             "long"            "easting_12n"
## [7] "northing_12n"    "buffer_dist"     "cfi_site"
## [10] "cfi_site_with_harvest" "cfi_site_with_vegedges" "harvest_0_15"
## [13] "harvest_gt_15"   "harvest_total"   "osm_industrial"
## [16] "pipe_trans"      "railways"        "roads"
```

## [19] "seismic"	"seismic_lines"	"seismic_lines_3D"
## [22] "trails"	"veg_edges"	"wells_active"
## [25] "wells_inactive"	"wells_total"	"pct_betu_pap"
## [28] "fire_0_15"	"fire_gt_15"	"pct_lari_lar"
## [31] "lc_broadleaf"	"lc_coniferous"	"lc_herbs"
## [34] "lc_mixedwood"	"lc_shrubs"	"lc_water"
## [37] "lc_wetland"	"lc_wetland_treed"	"pct_pice_gla"
## [40] "pct_pice_mar"	"pct_pinu_ban"	"pct_popu_tre"
## [43] "nonanthro_cai_mn"	"nonanthro_ed"	"nonanthro_tca"
## [46] "landscape_cai_mn"	"landscape_ed"	"landscape_tca"
## [49] "nonanthro_cohesion"	"landscape_cohesion"	"landscape_contag"
## [52] "landscape_mesh"	"landscape_np"	"landscape_shei"
## [55] "landscape_siei"		

Identify the covariates of interest.

```
covs_to_explore <- covs %>%

# To keep the number of plots reasonable we'll remove some things we're 99% sure we don't want to
#   model
select(
  -contains("gt"),
  -railways,
  -seismic_lines,
  -seismic_lines_3D,
  -trails,
  -veg_edges,
  -lc_water,
  -pct_betu_pap,
  -landscape_cohesion,
  -landscape_ed,
  -landscape_cai_mn,
  -landscape_tca) %>%

relocate((1:8), sort(names(.)))
```

## 2. Data distribution

### 2.1. Histograms at 500 meters:

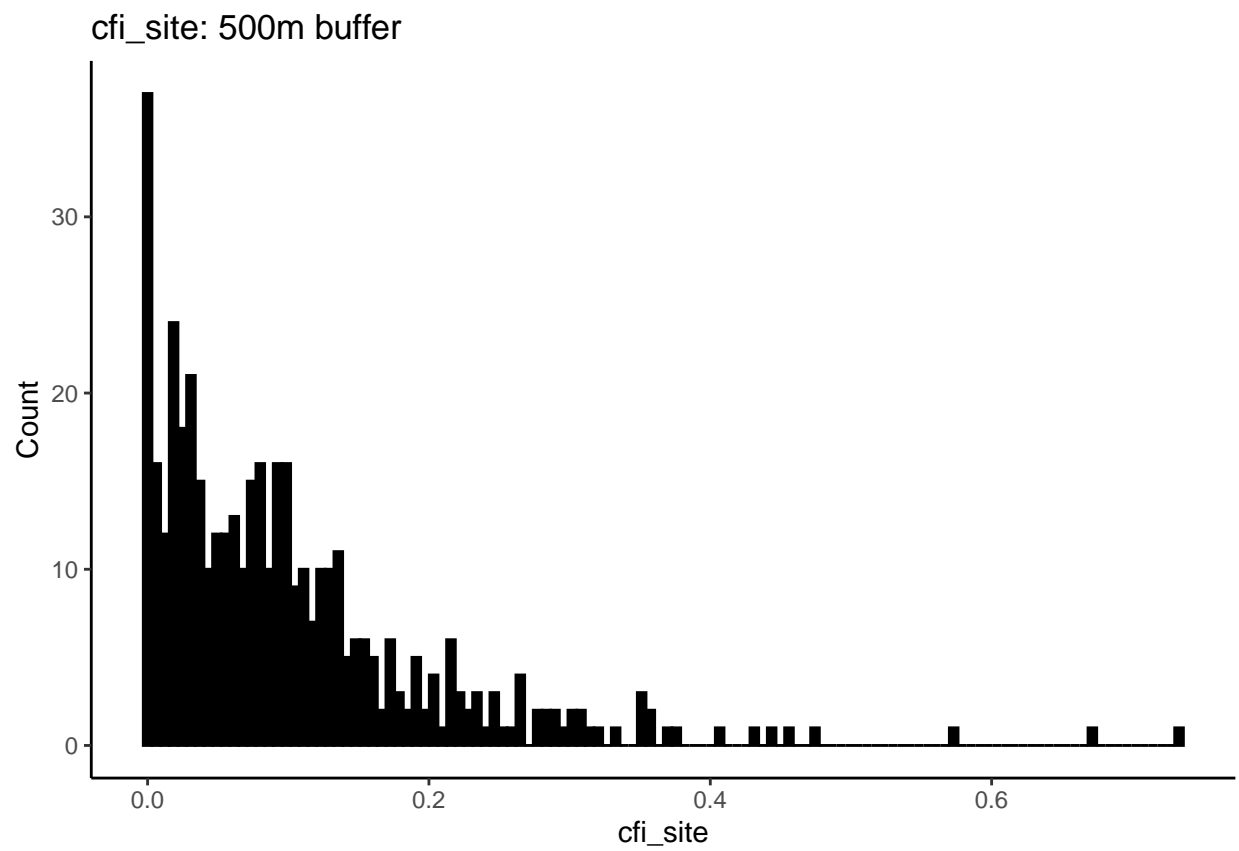
```
covs_to_explore %>%

filter(buffer_dist == 500) %>%

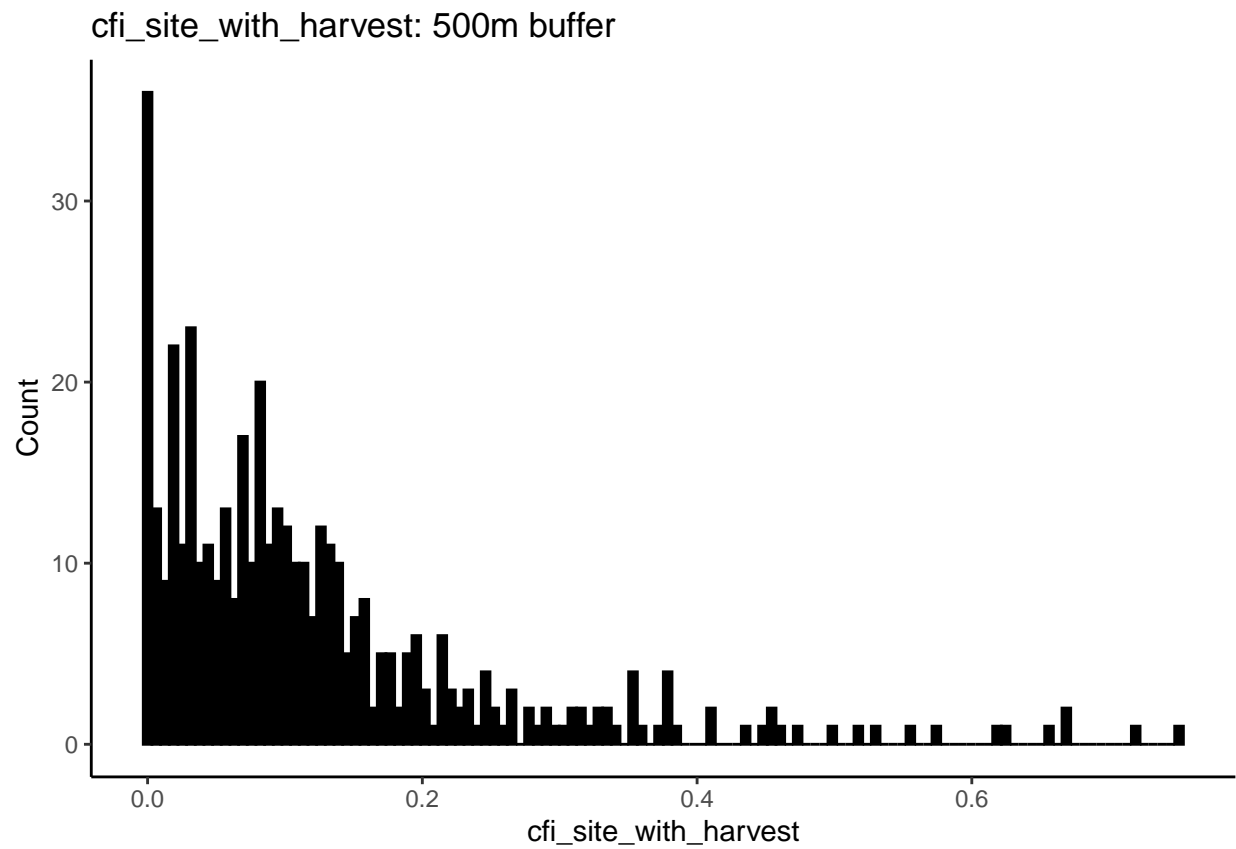
# Remove the response columns
select(-(1:8)) %>%

# use imap which will retain both the data (x) and the variable names (y)
imap(~ ggplot(mapping = aes(x = .x)) +
  geom_histogram(bins = 120, fill = "black", color = "black") +
  labs(title = paste0(.y, ": 500m buffer"), x = .y, y = "Count") +
  theme_classic())
```

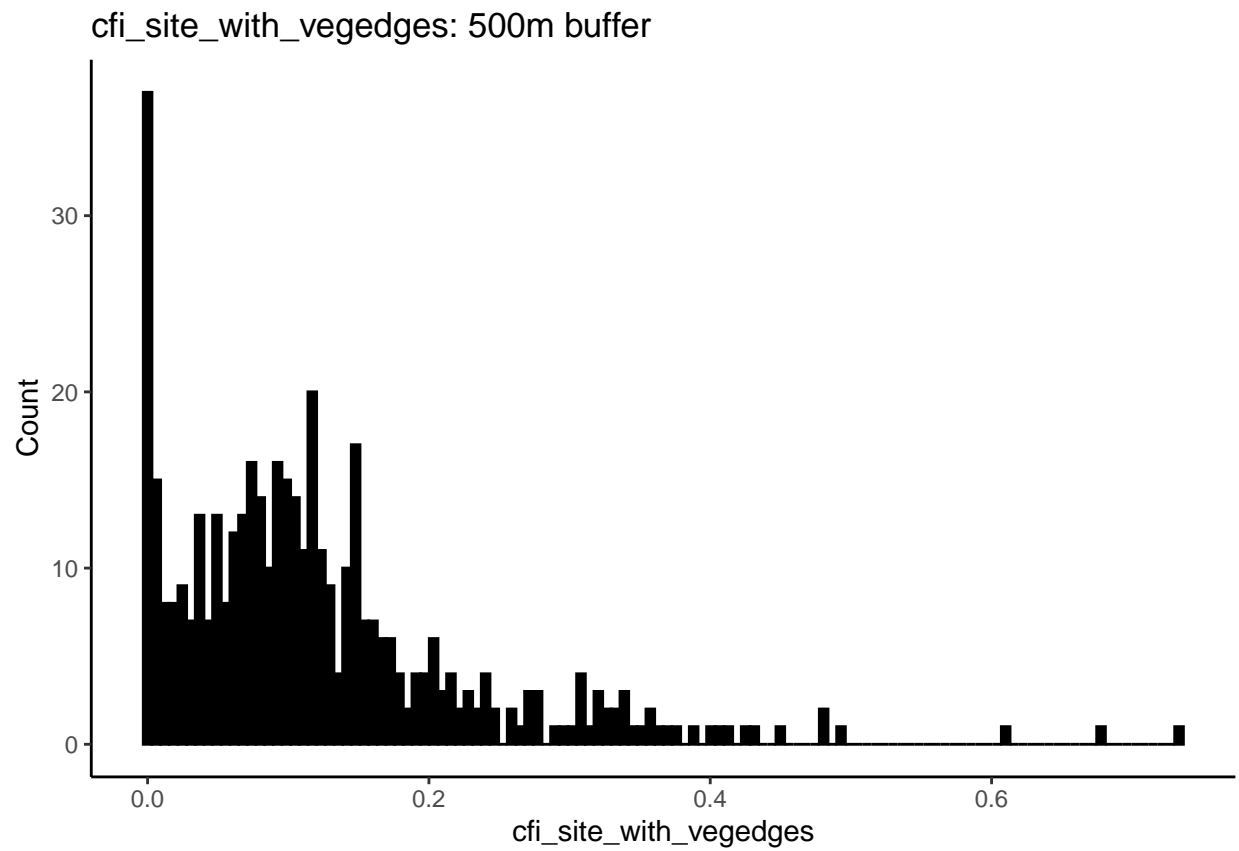
```
## $cfi_site
```



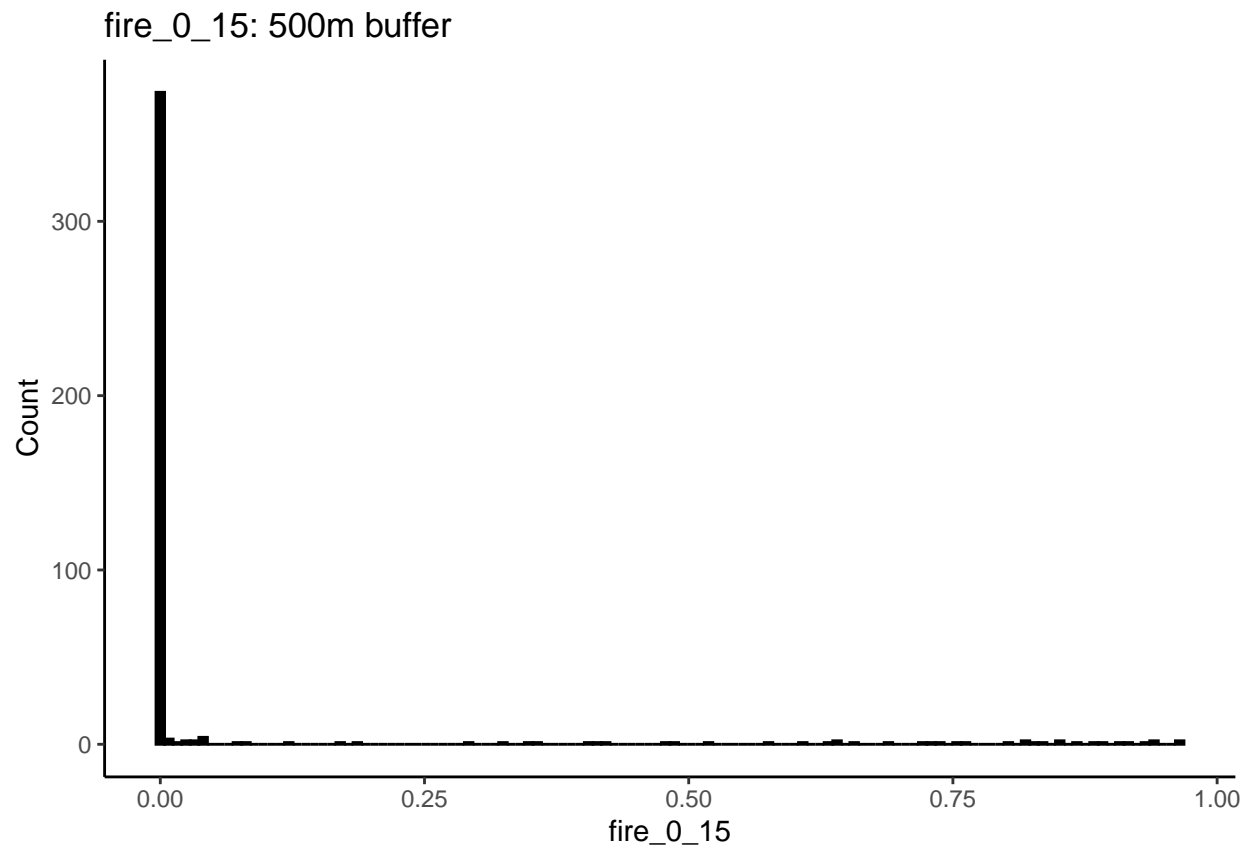
```
##  
## $cfi_site_with_harvest
```



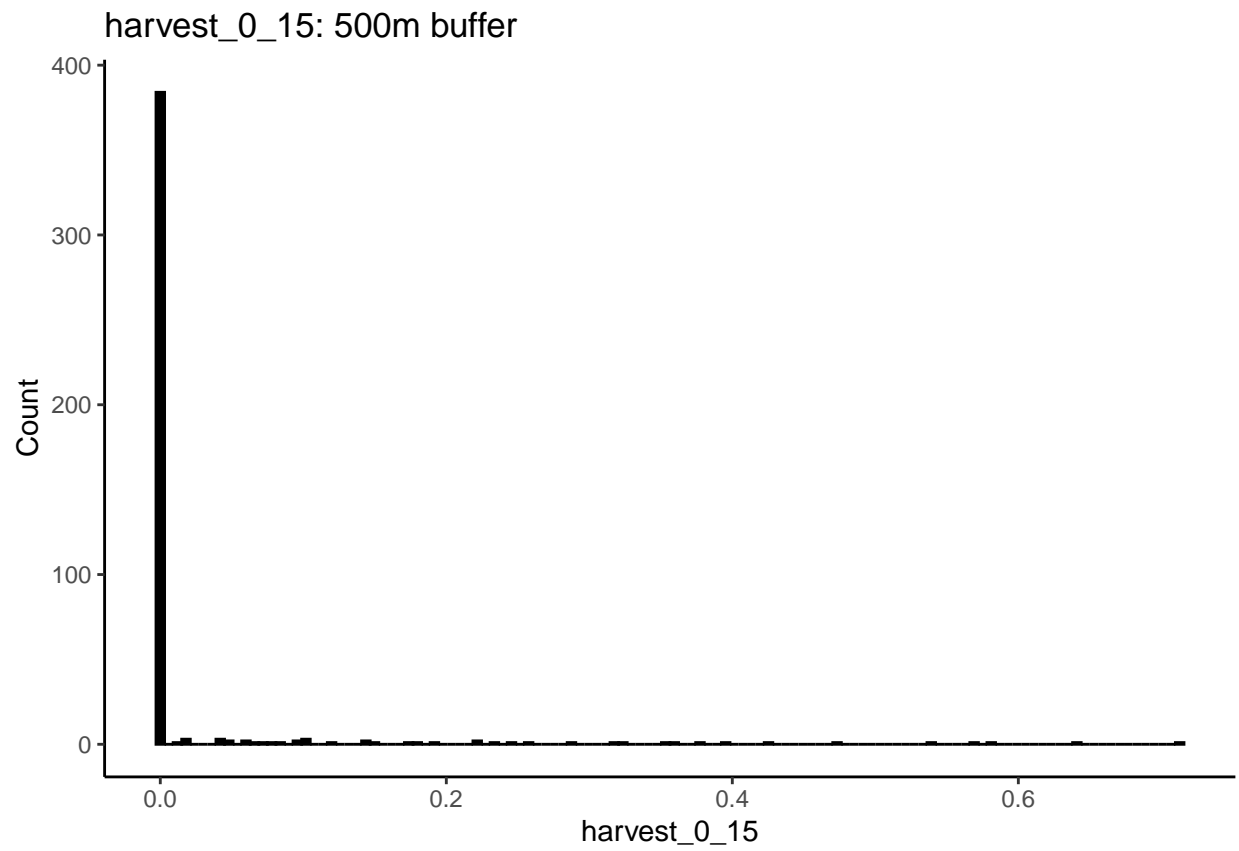
```
##  
## $cfi_site_with_vegedges
```



```
##  
## $fire_0_15
```

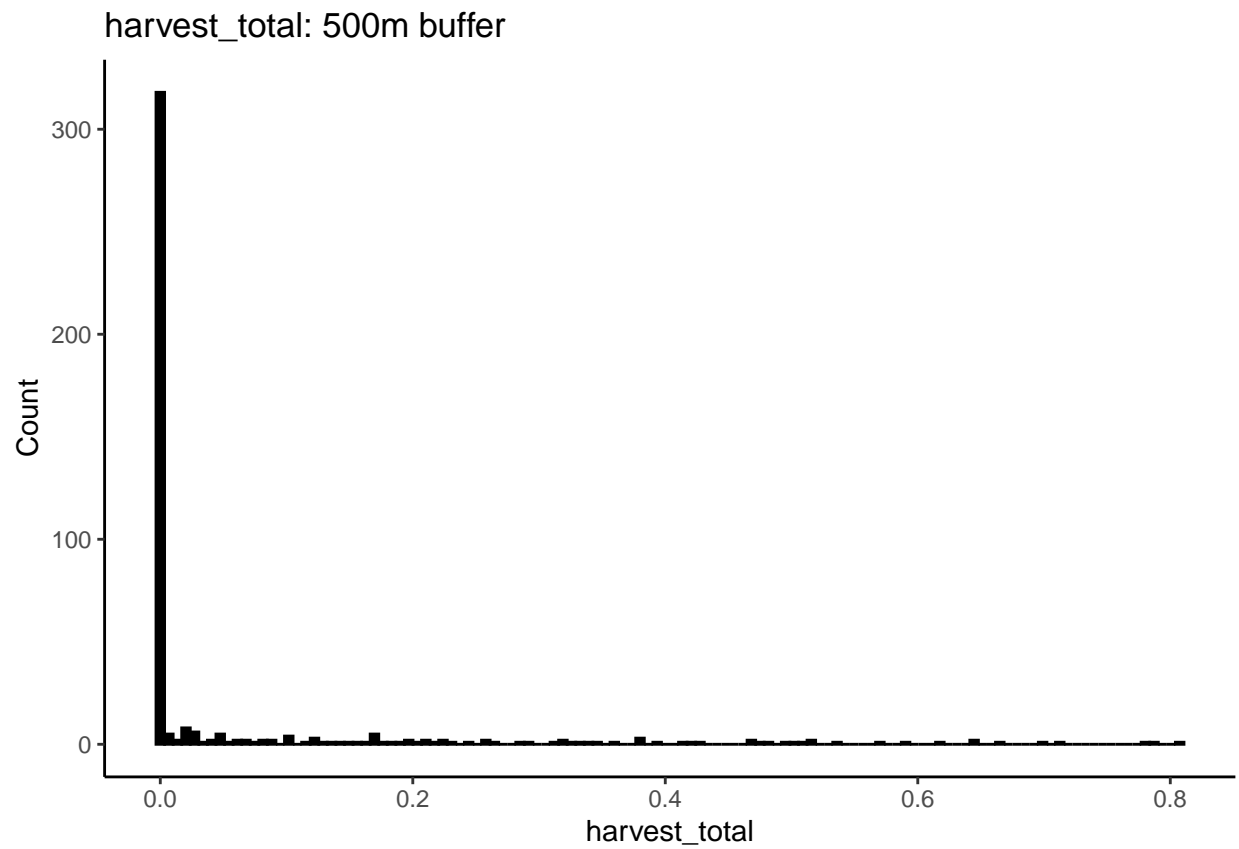


```
##  
## $harvest_0_15
```

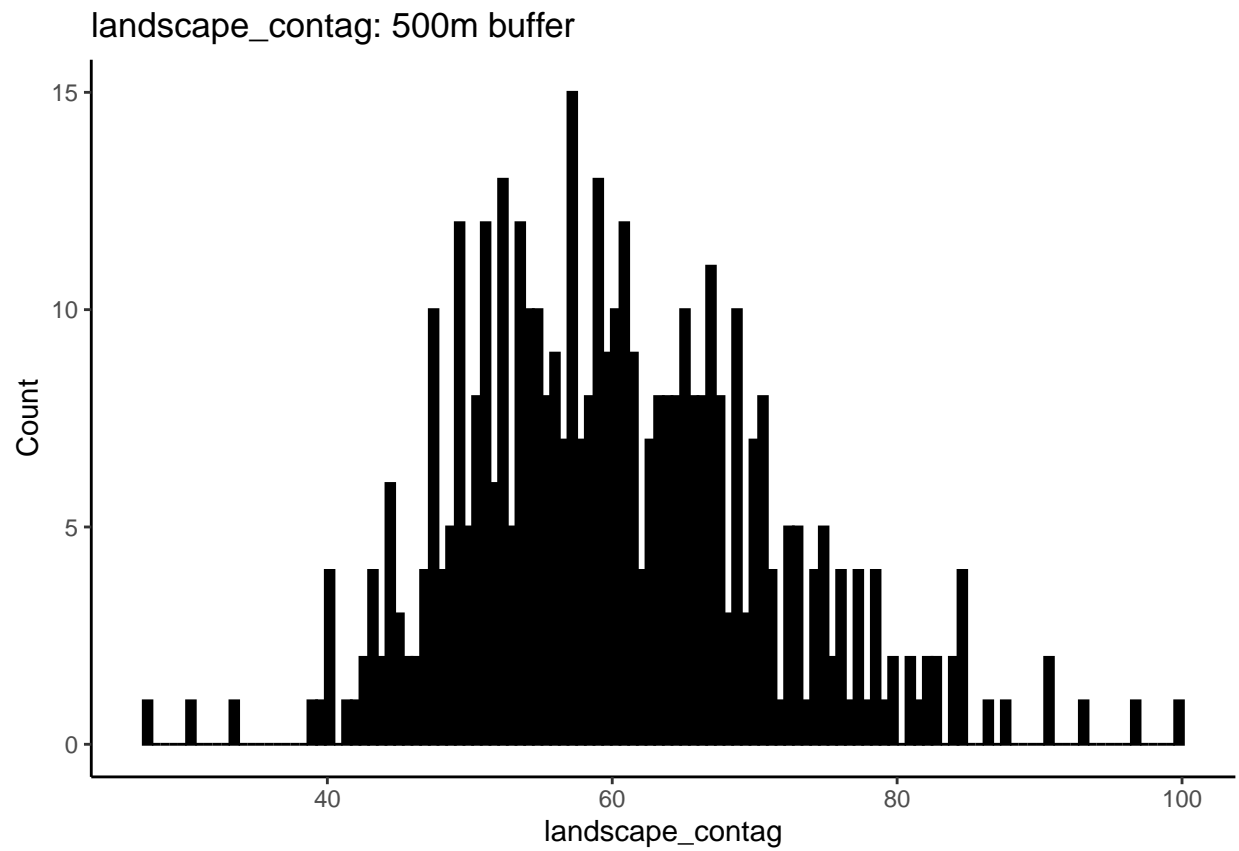


```
##  
## $harvest_total
```

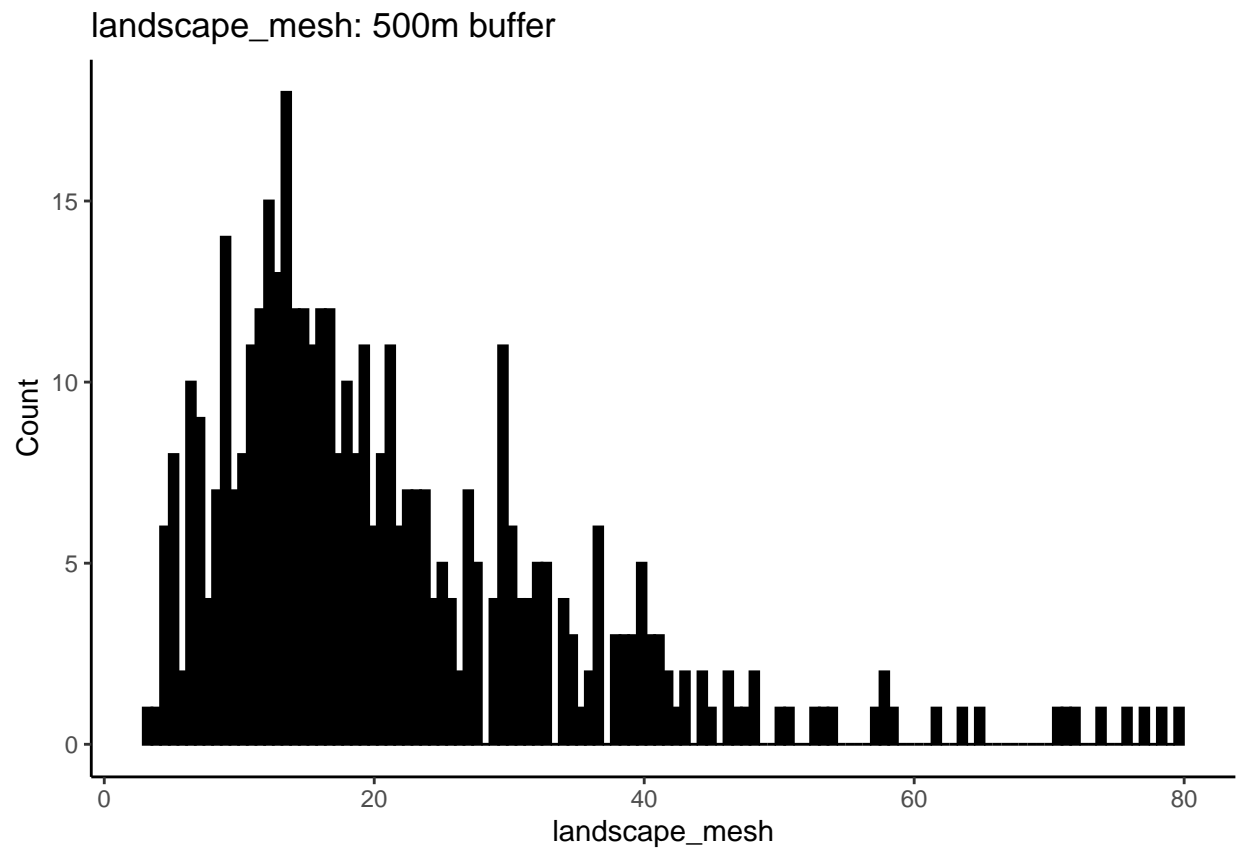




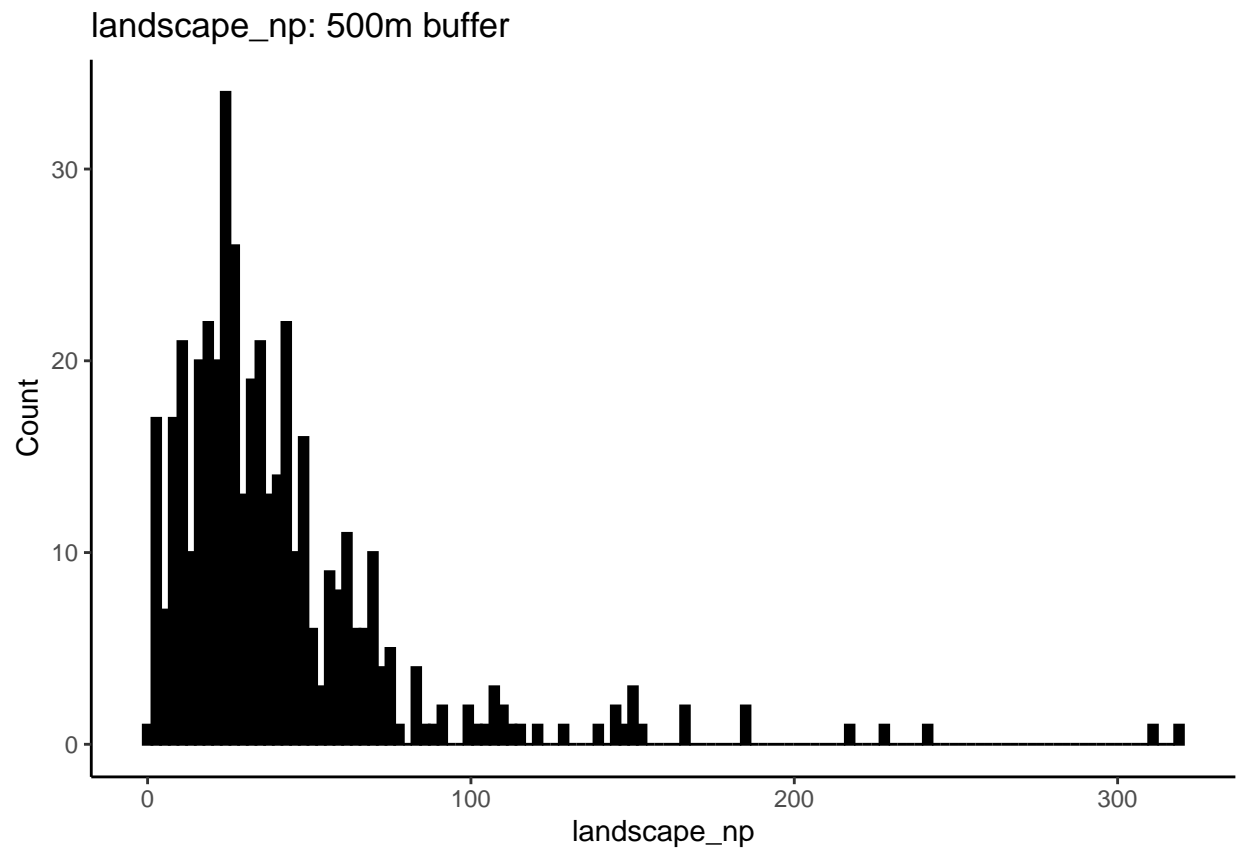
```
##  
## $landscape_contag
```



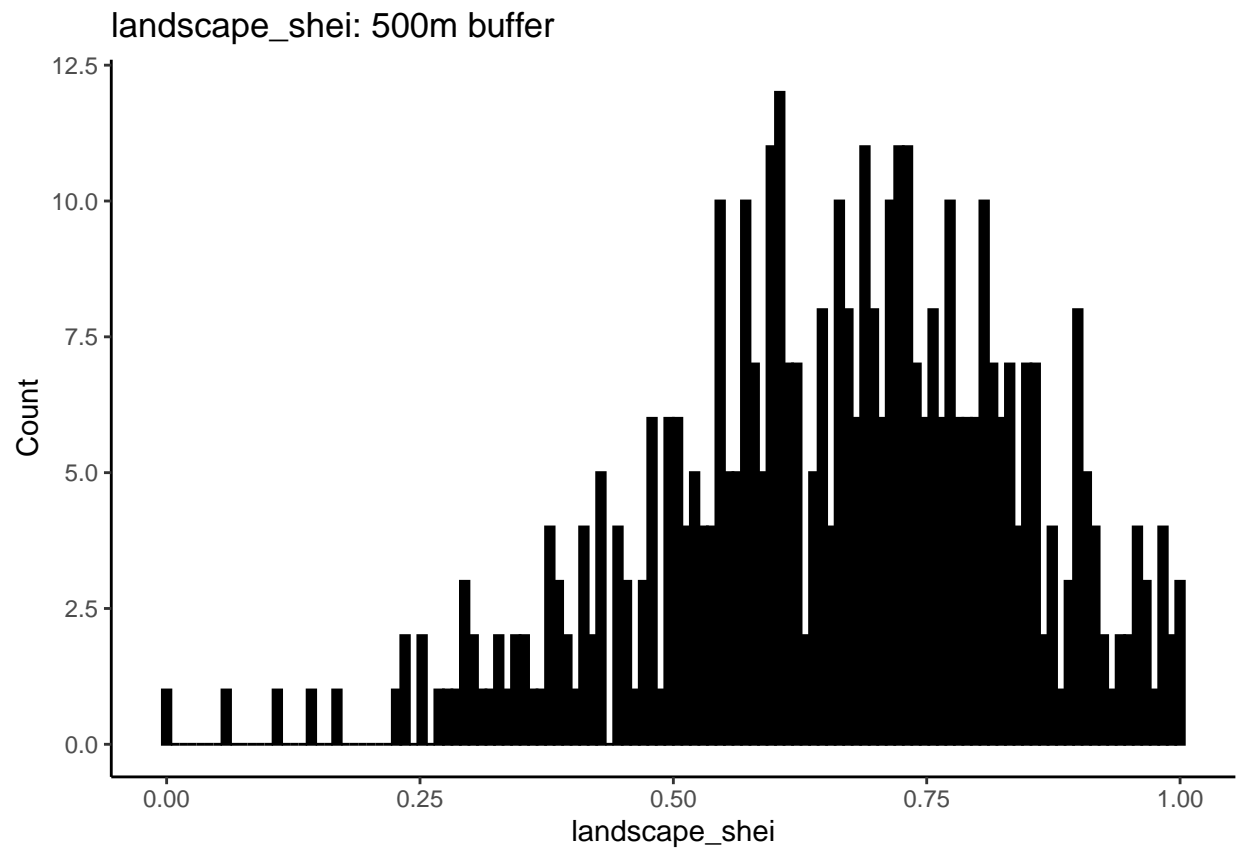
```
##  
## $landscape_mesh
```



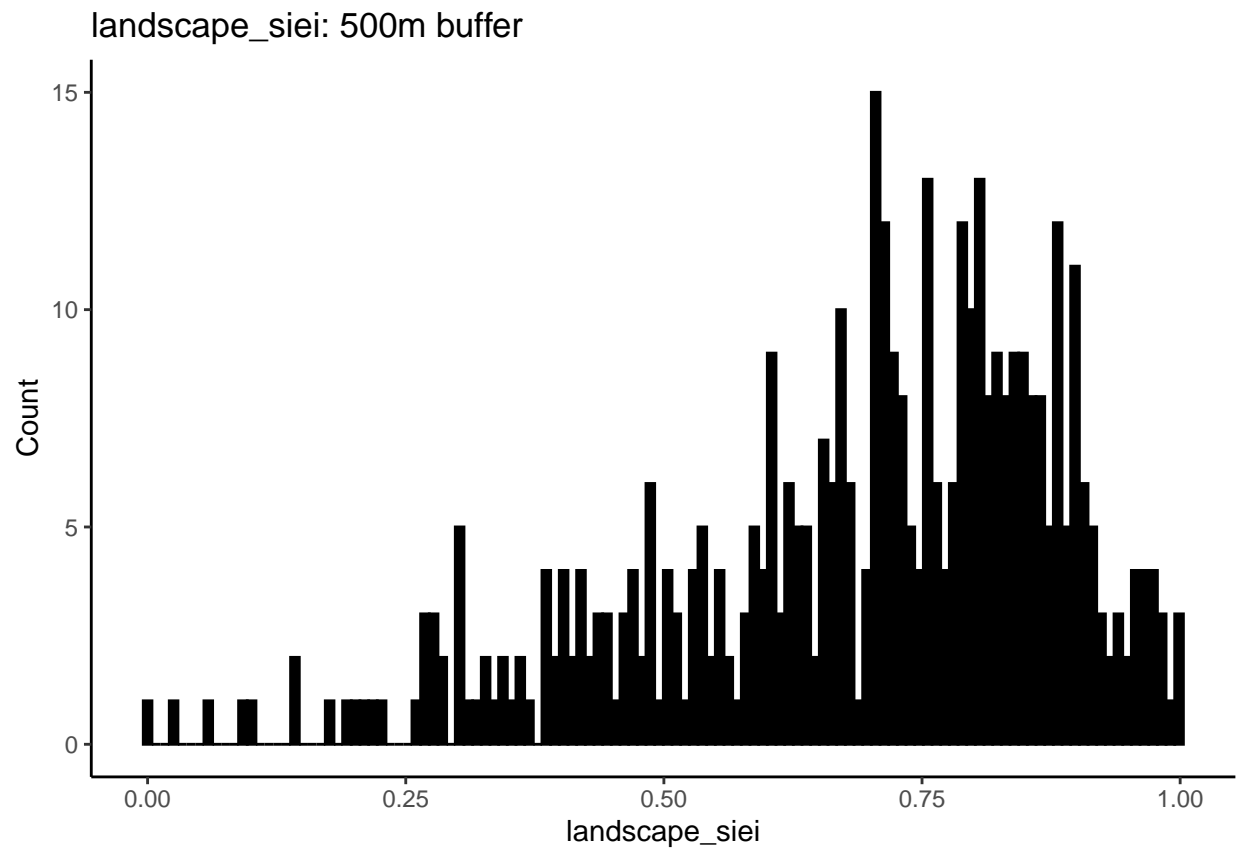
```
##  
## $landscape_np
```



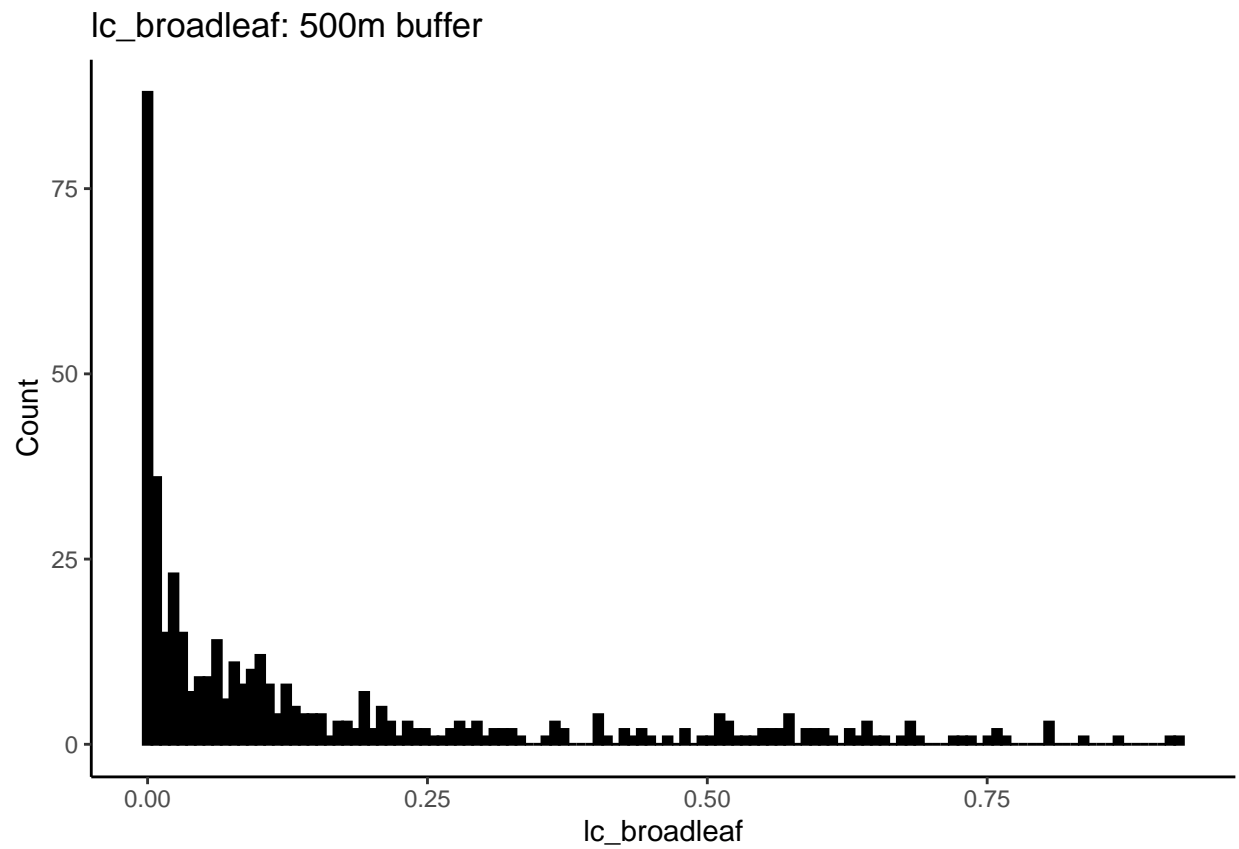
```
##  
## $landscape_shei
```



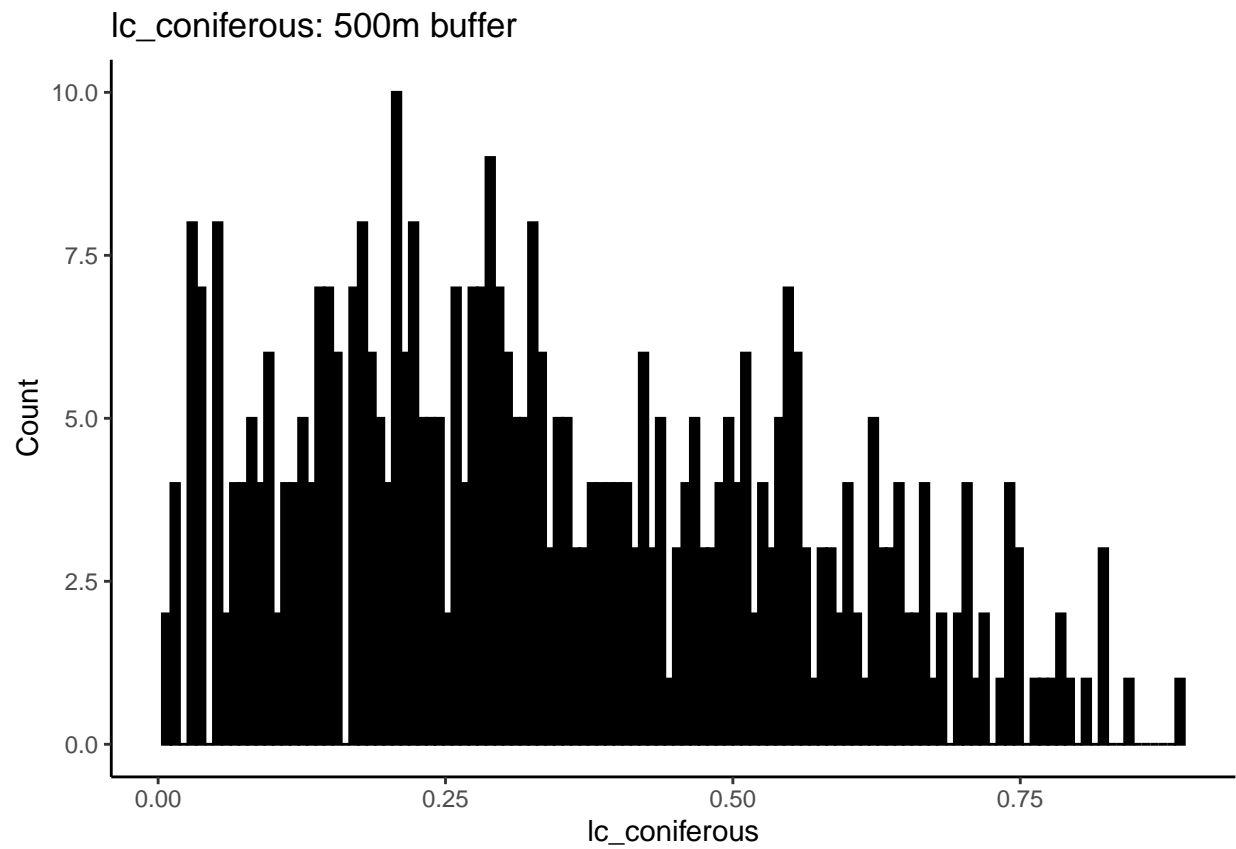
```
##  
## $landscape_siei
```



```
##  
## $lc_broadleaf
```

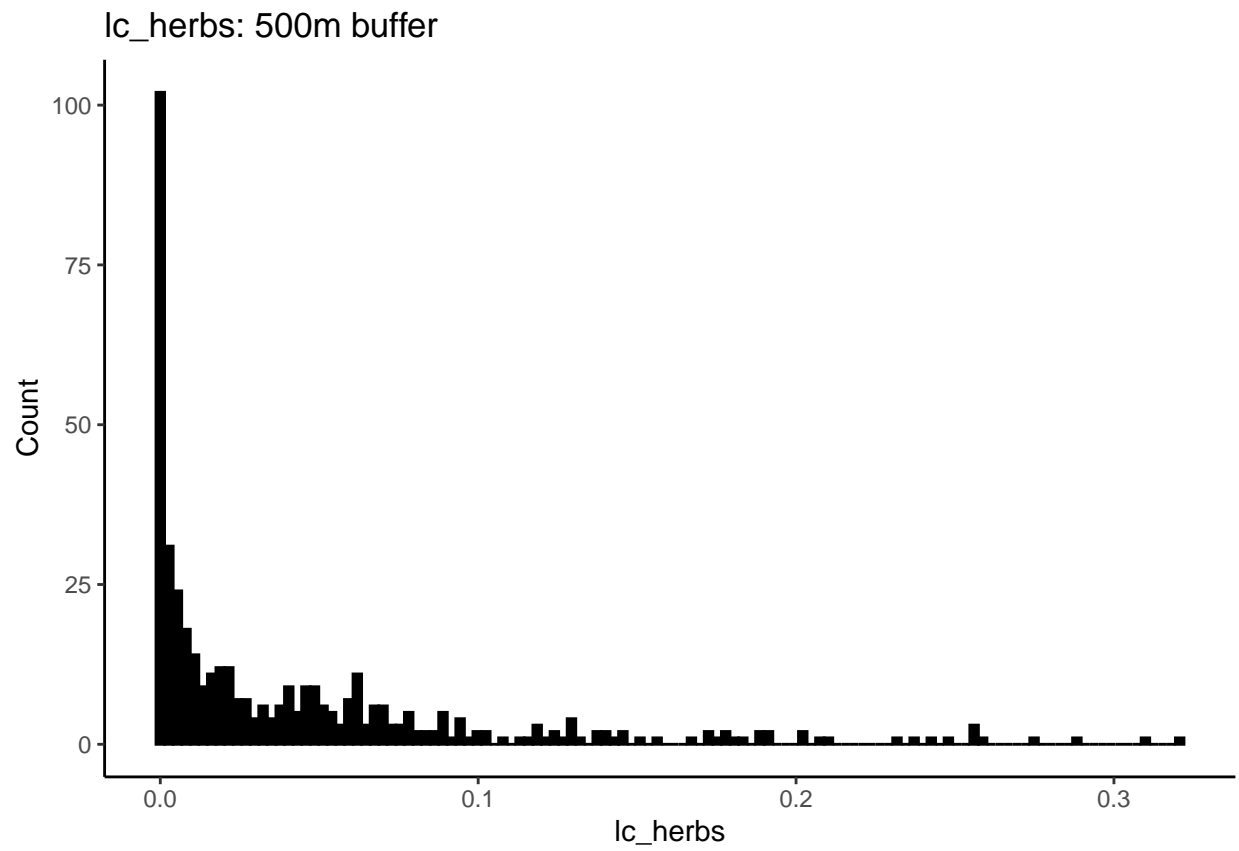


```
##  
## $lc_coniferous
```

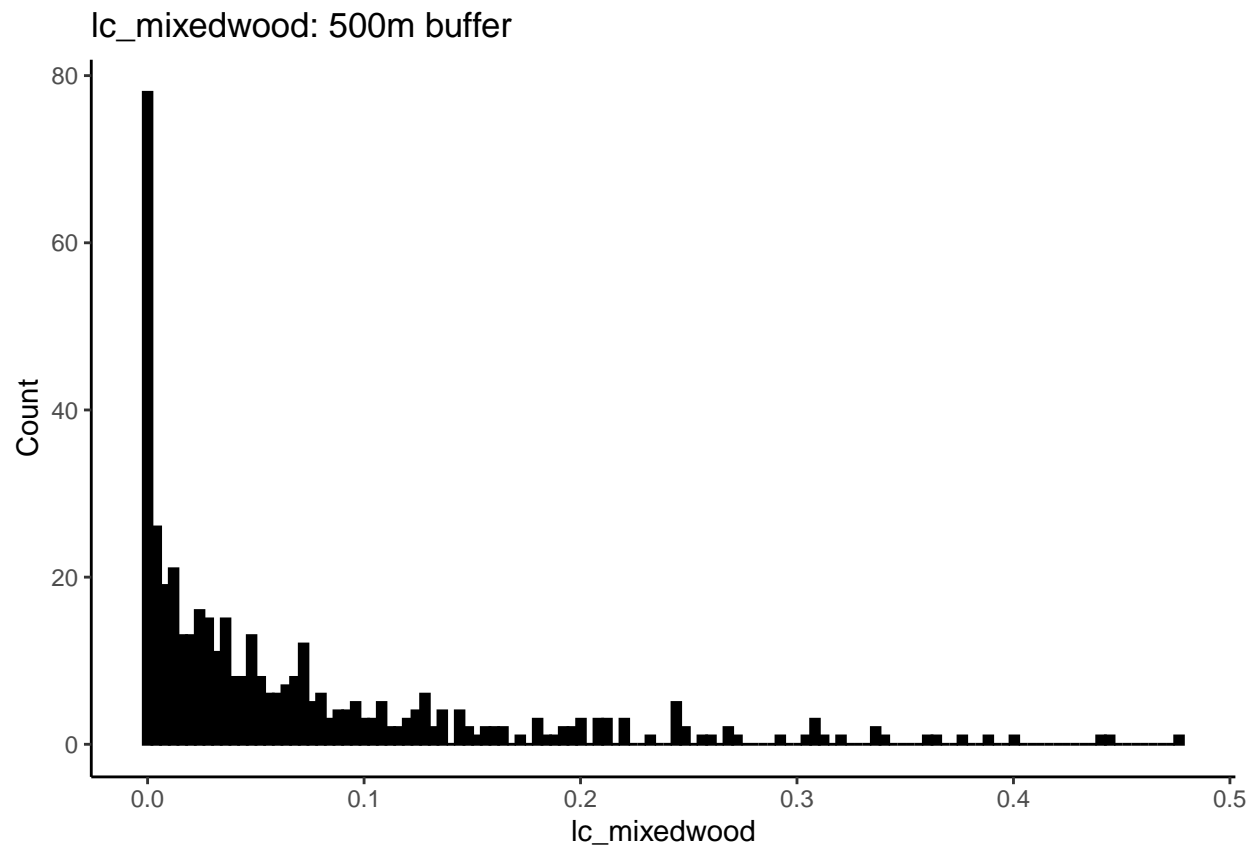


```
##  
## $lc_herbs
```

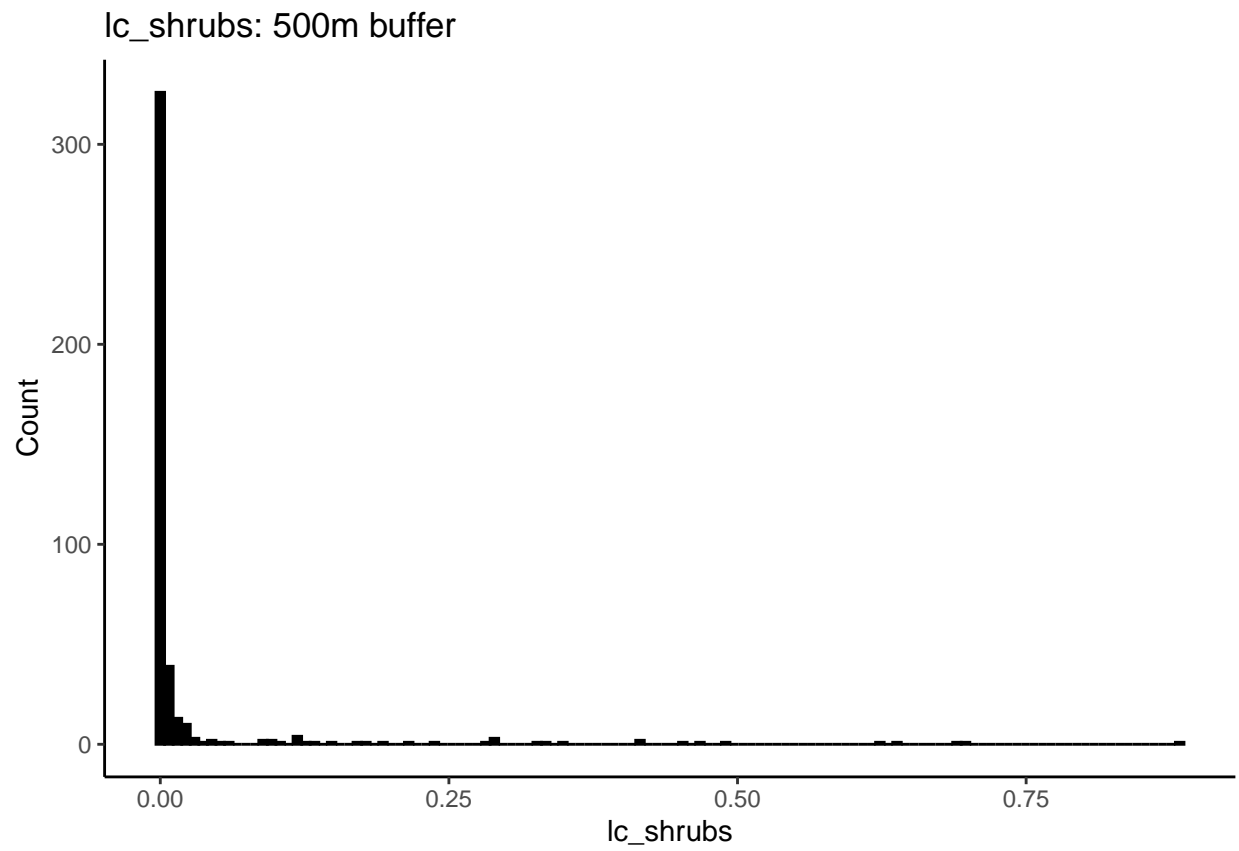




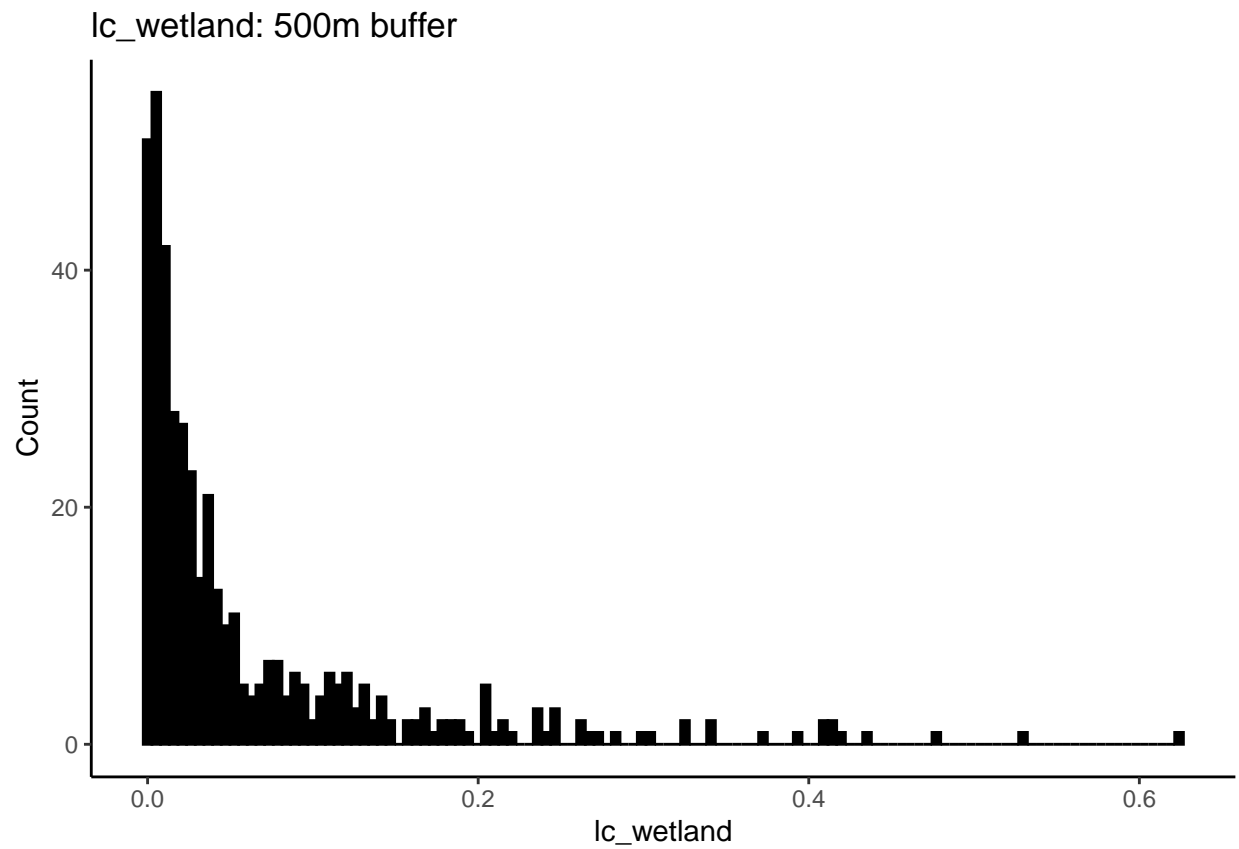
```
##  
## $lc_mixedwood
```



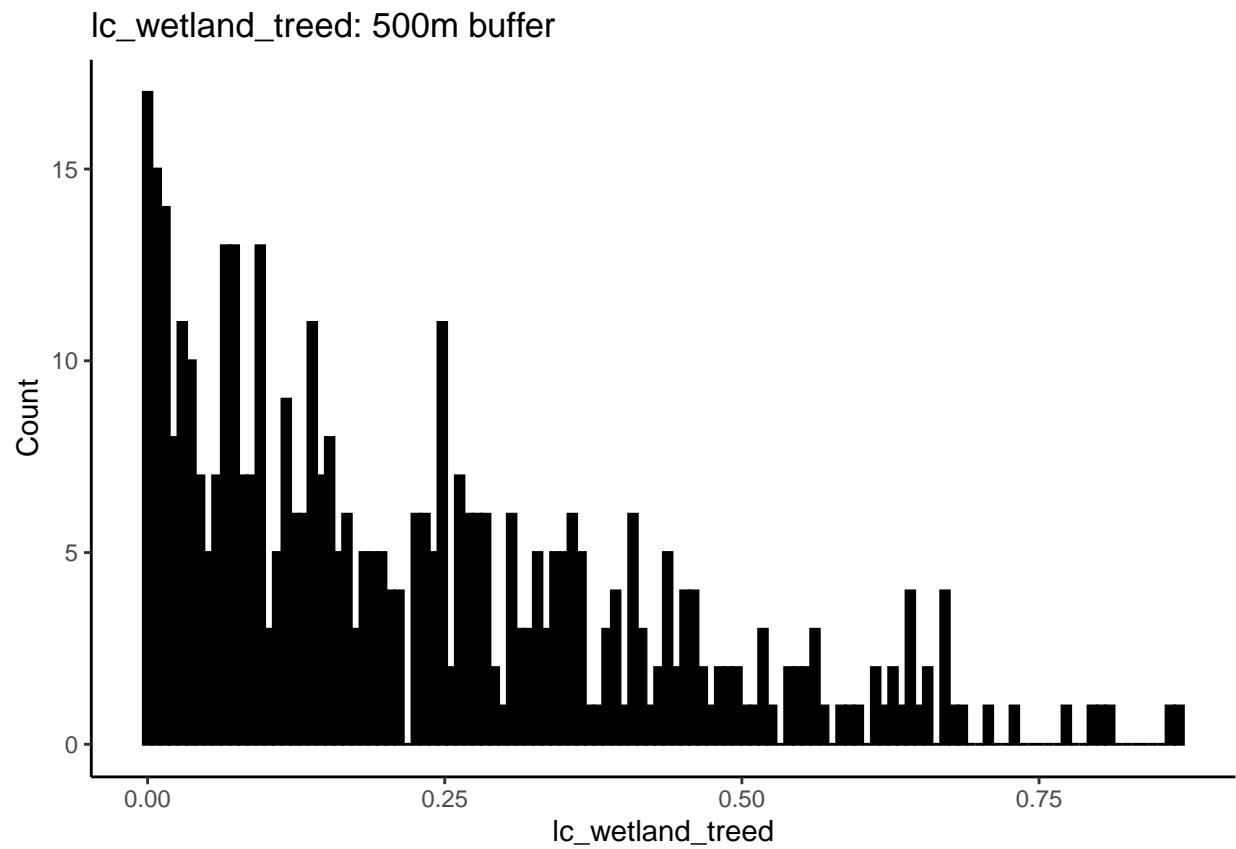
```
##  
## $lc_shrubs
```



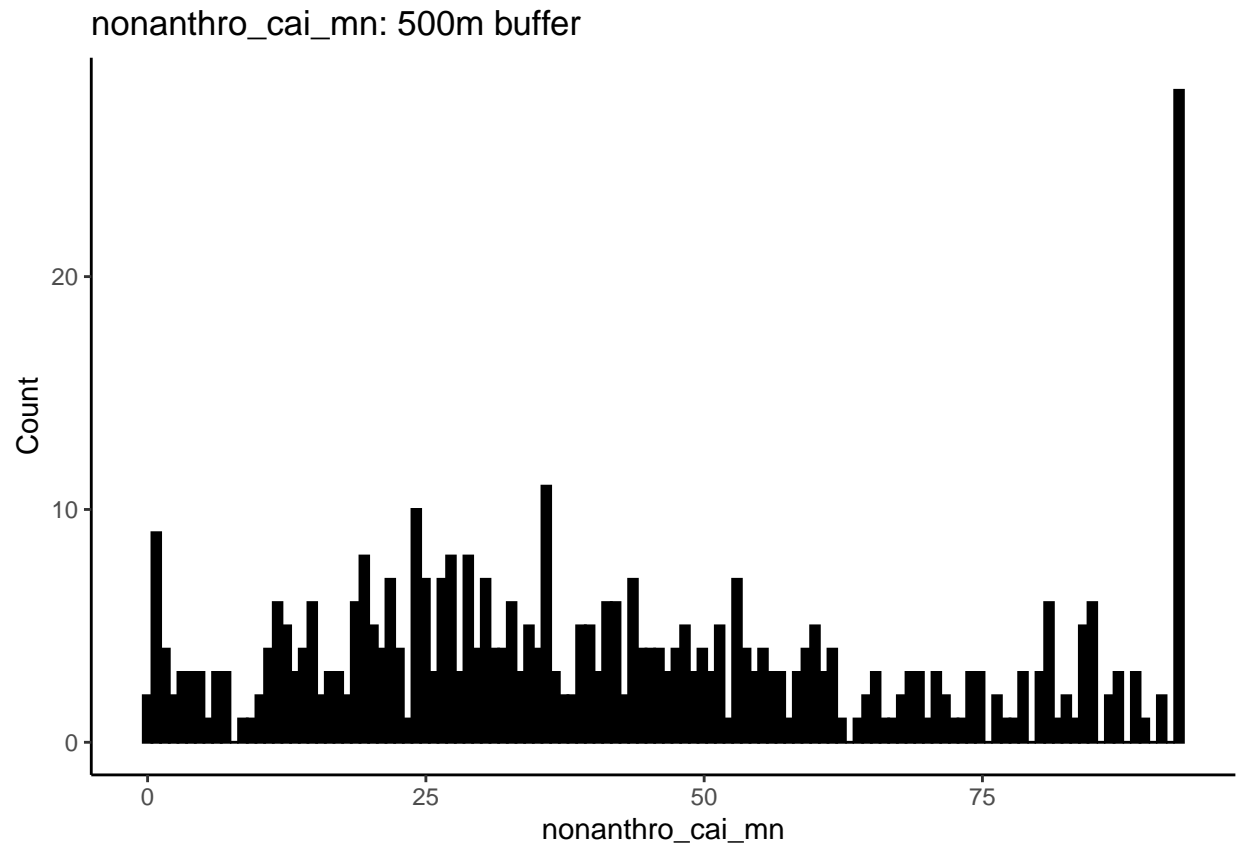
```
##  
## $lc_wetland
```



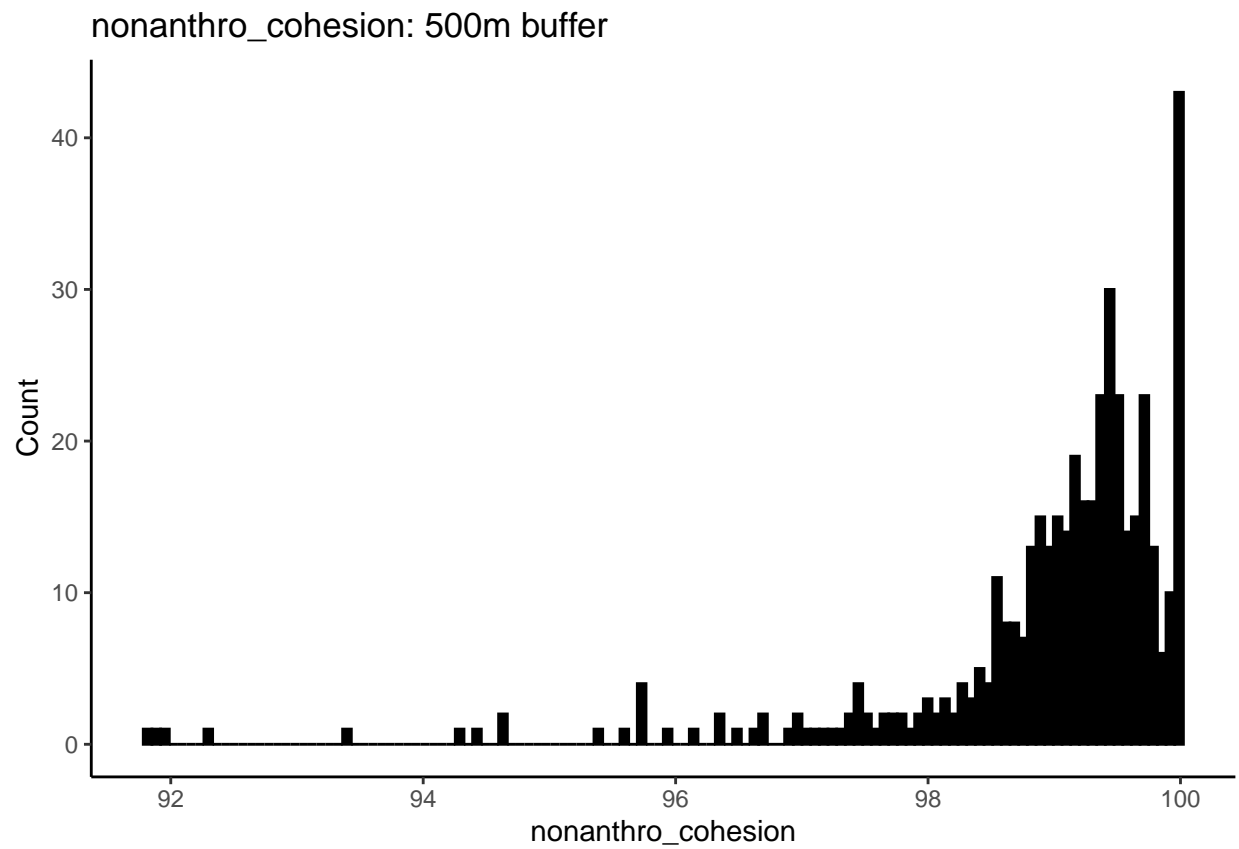
```
##  
## $lc_wetland_treed
```



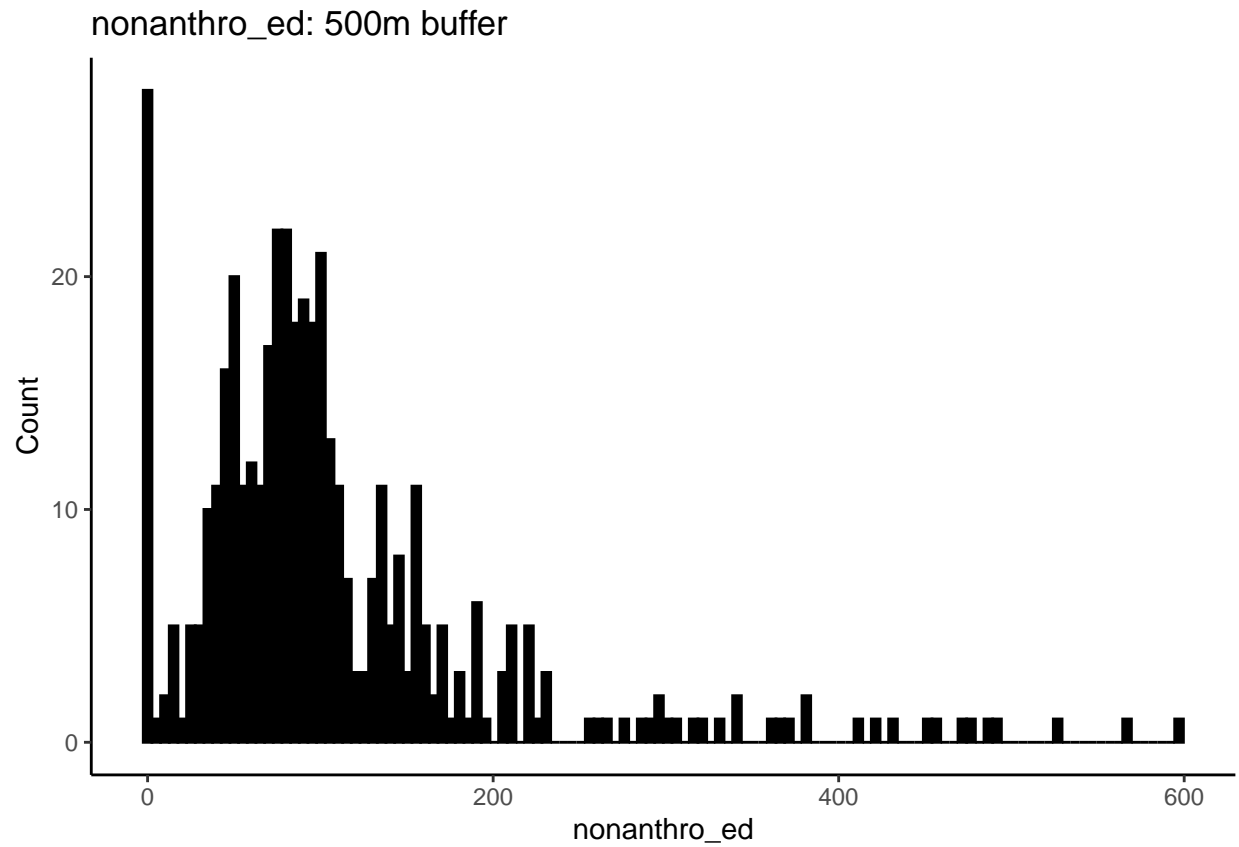
```
##  
## $nonanthro_cai_mn
```



```
##  
## $nonanthro_cohesion
```

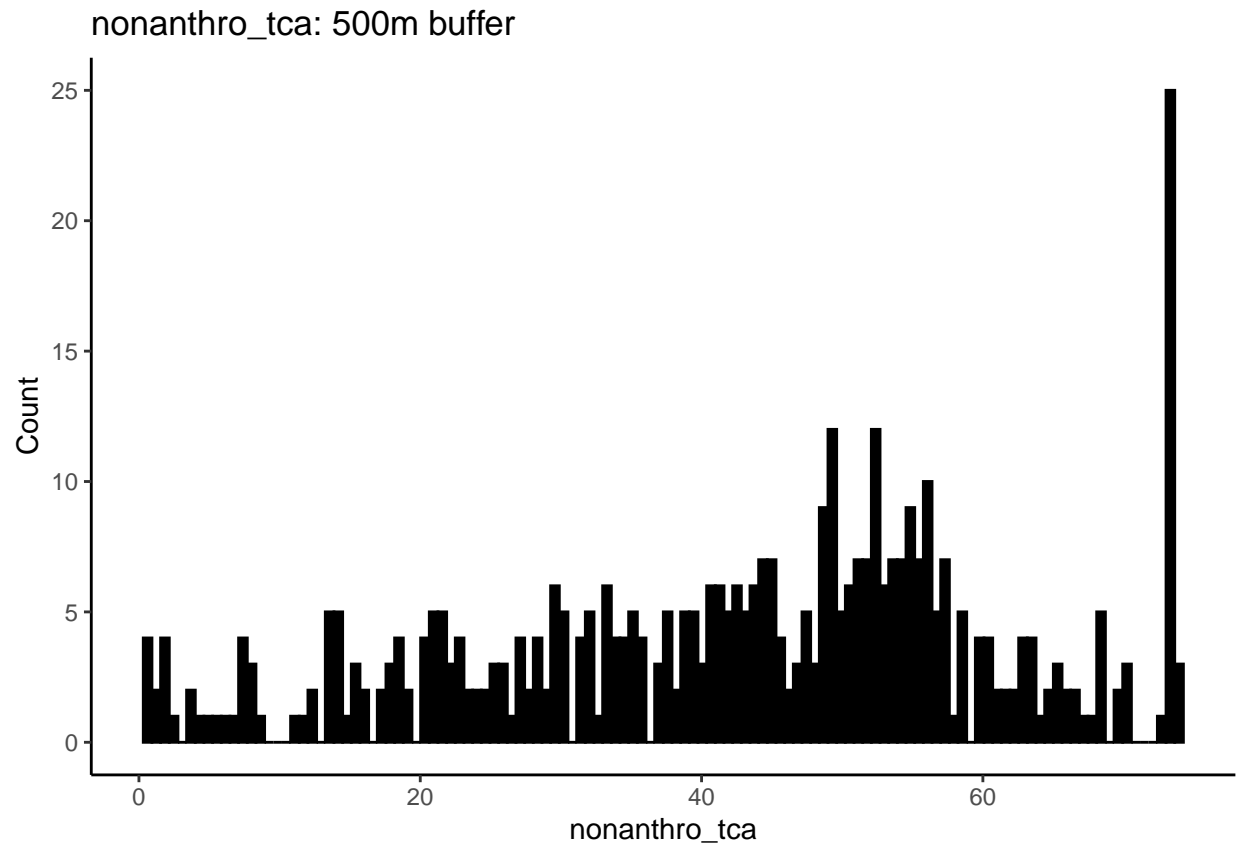


```
##  
## $nonanthro_ed
```

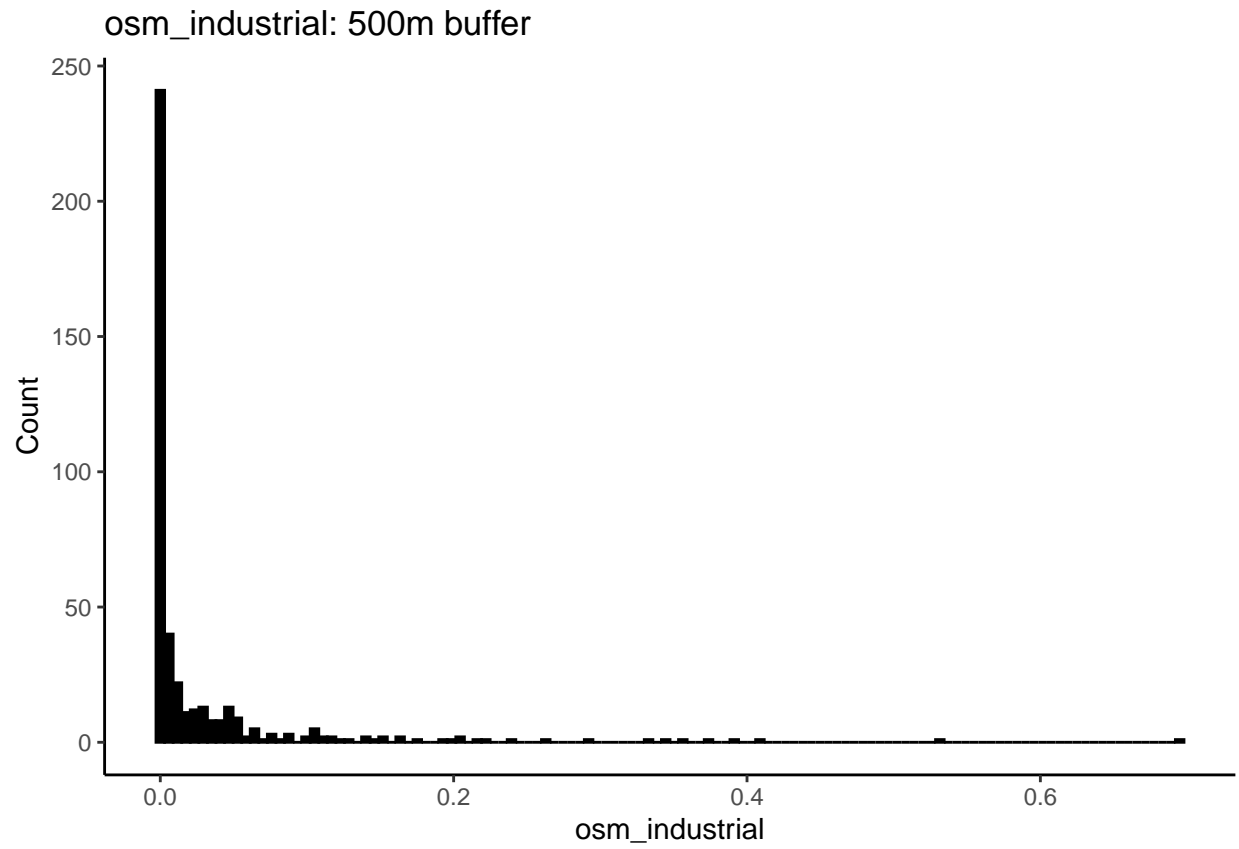


```
##  
## $nonanthro_tca
```

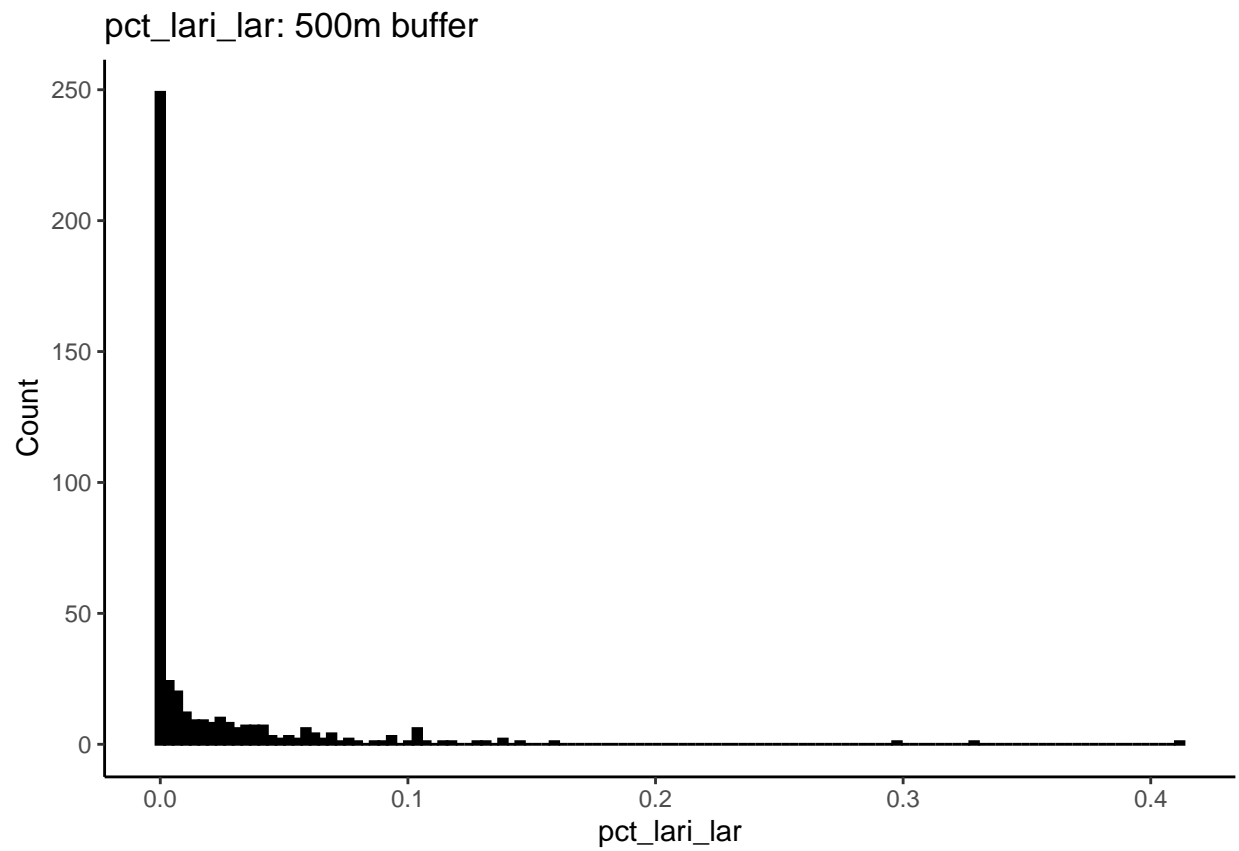




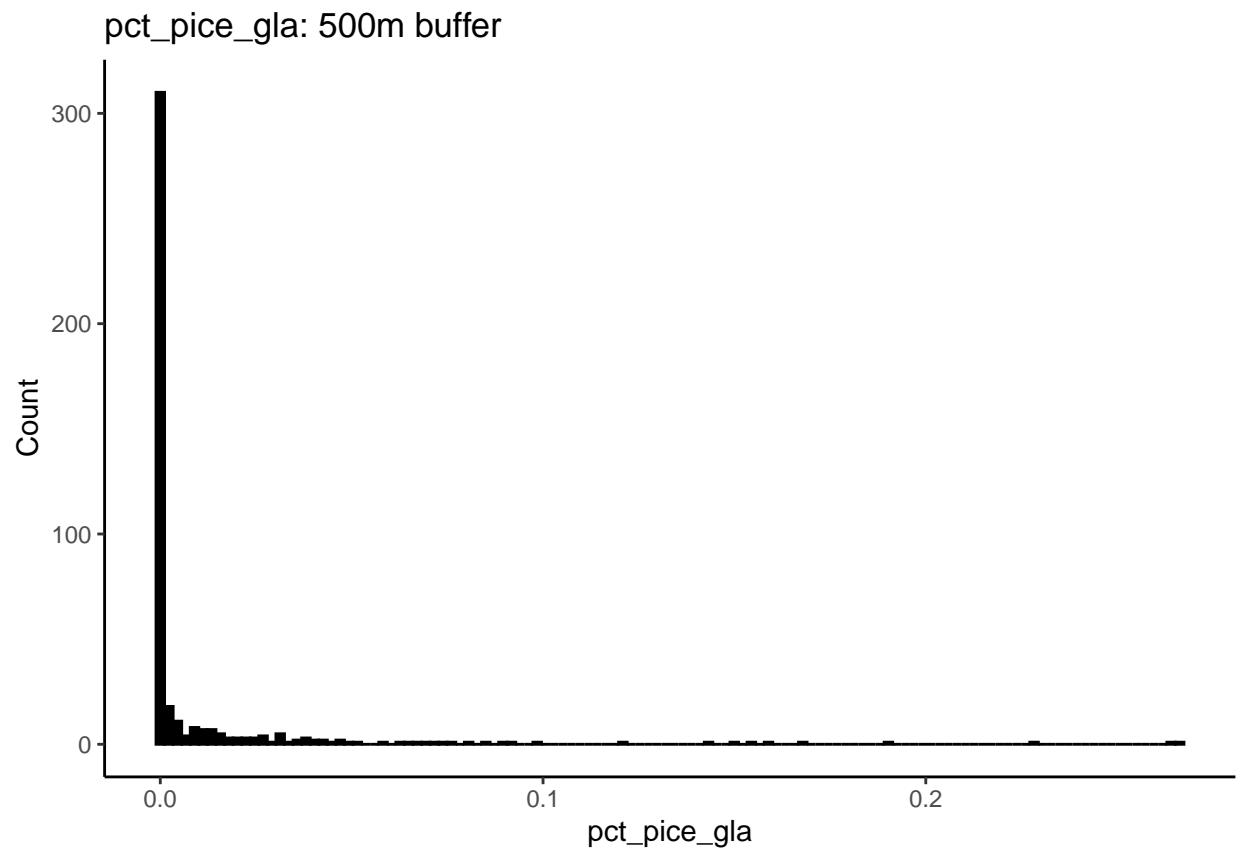
```
##  
## $osm_industrial
```



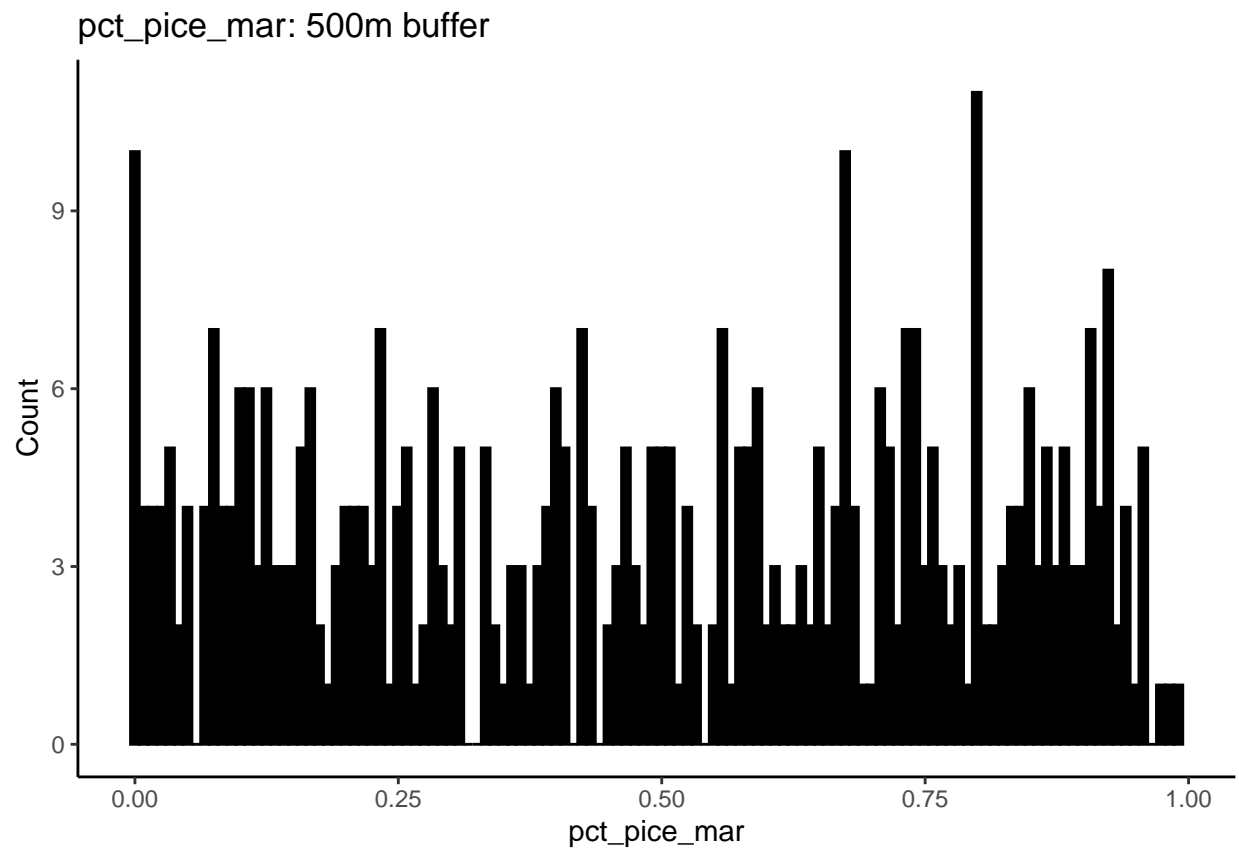
```
##  
## $pct_lari_lar
```



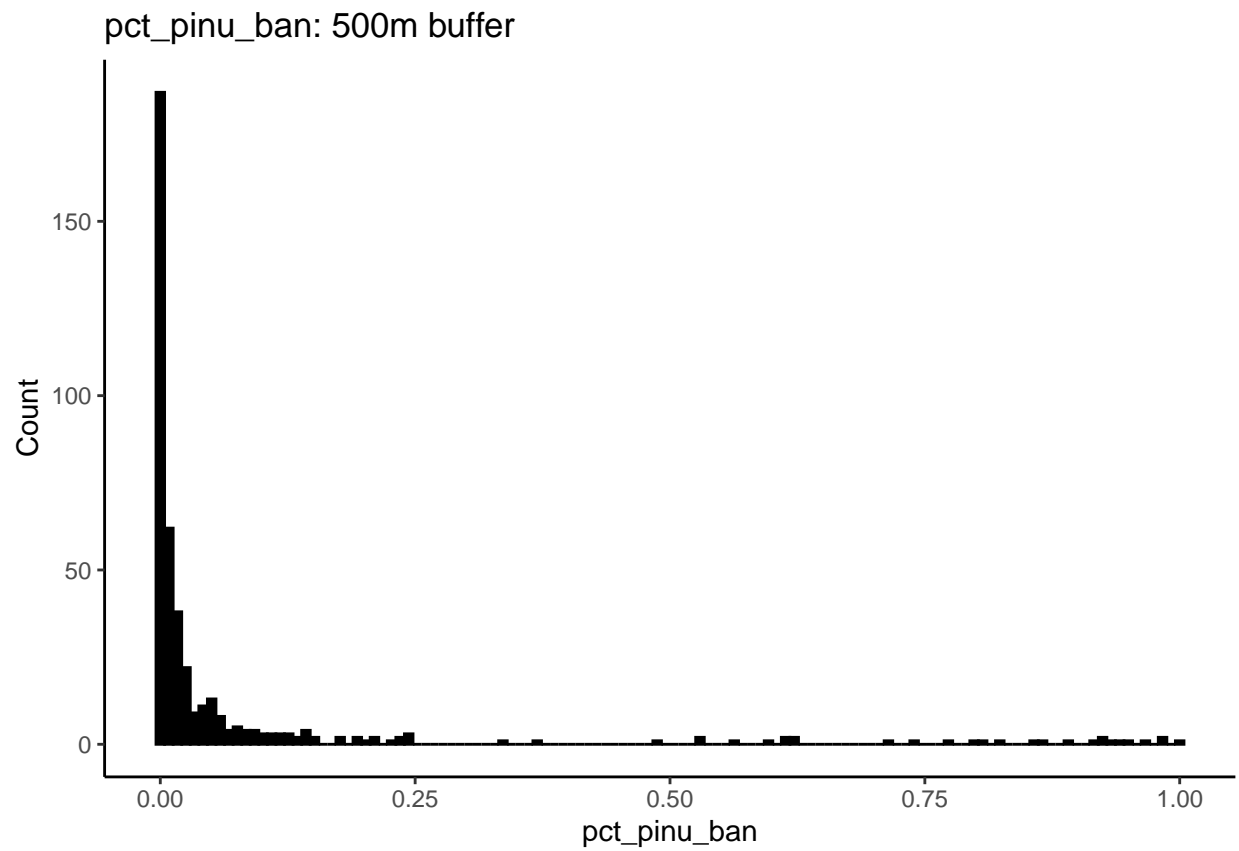
```
##  
## $pct_pice_gla
```



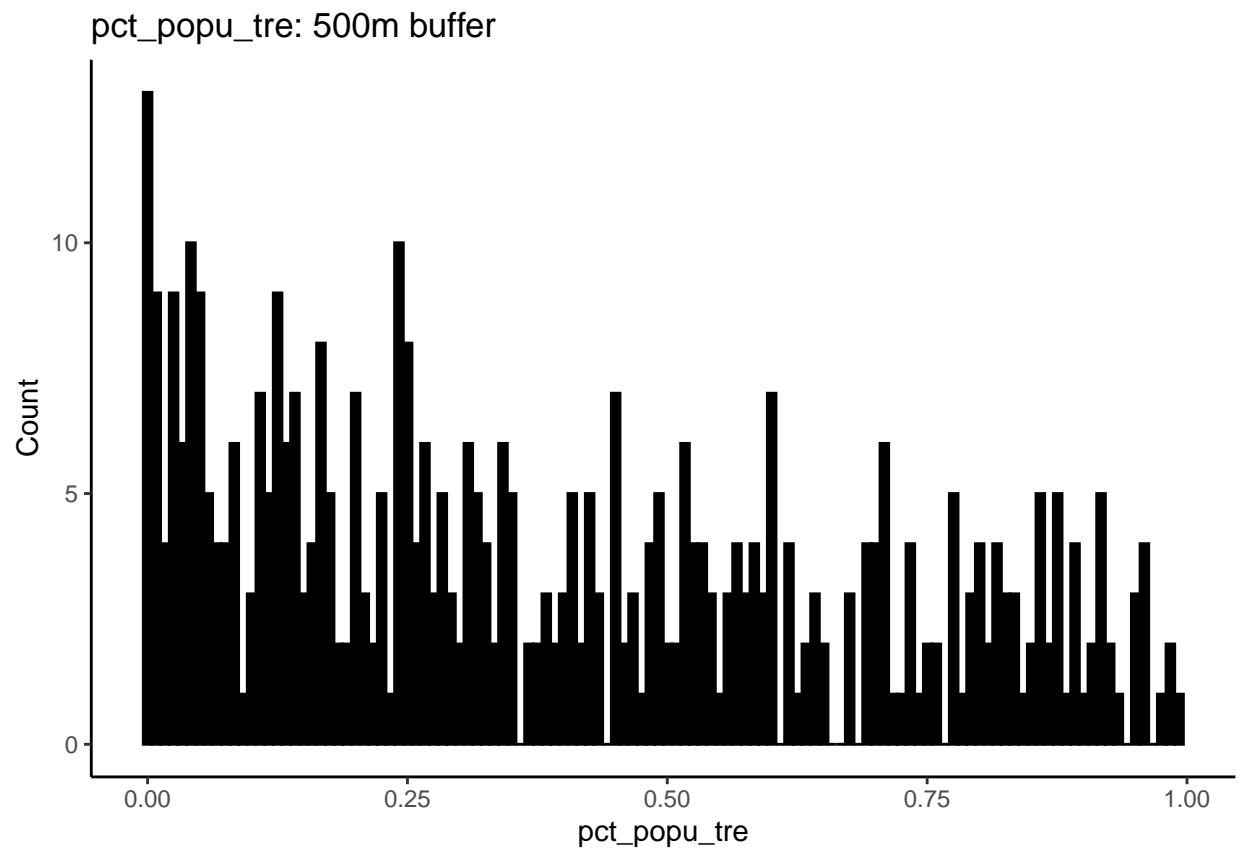
```
##  
## $pct_pice_mar
```



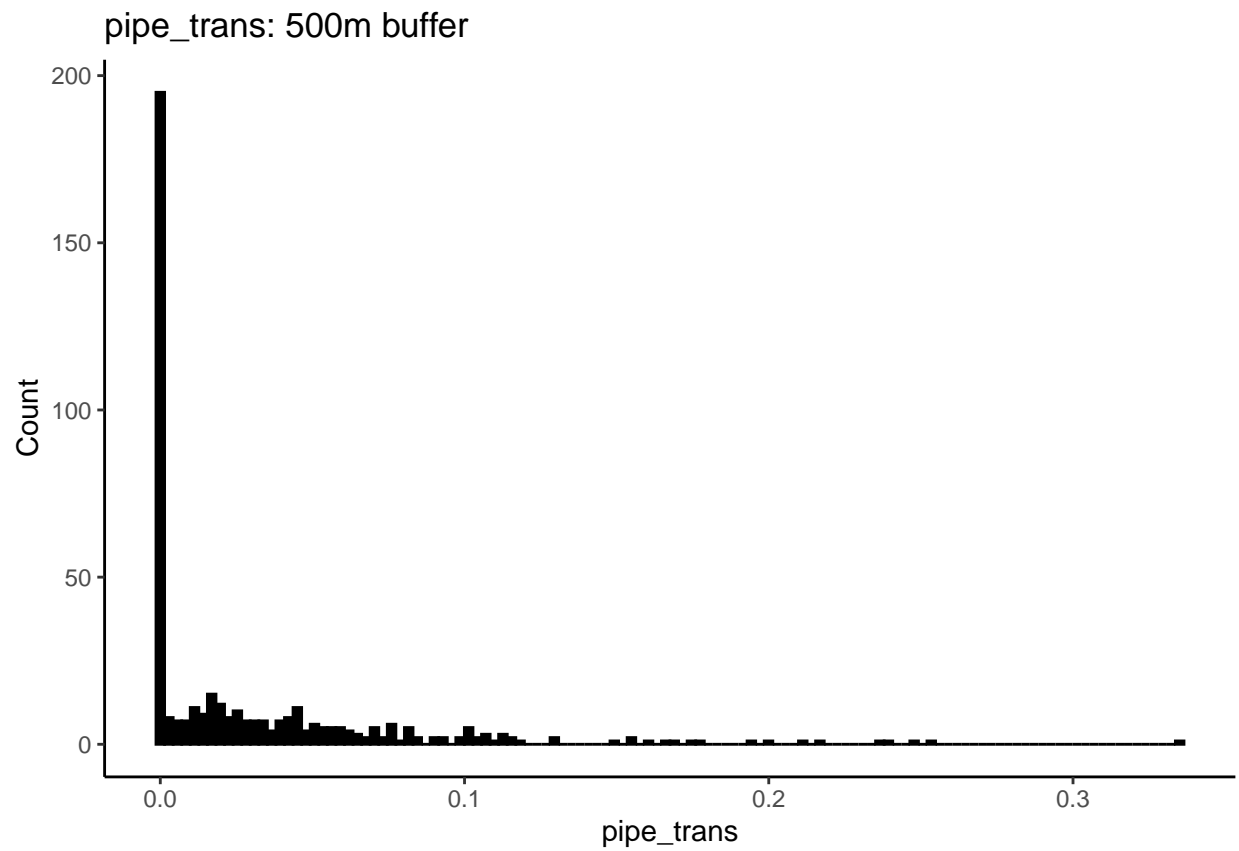
```
##  
## $pct_pinu_ban
```



```
##  
## $pct_popu_tre
```

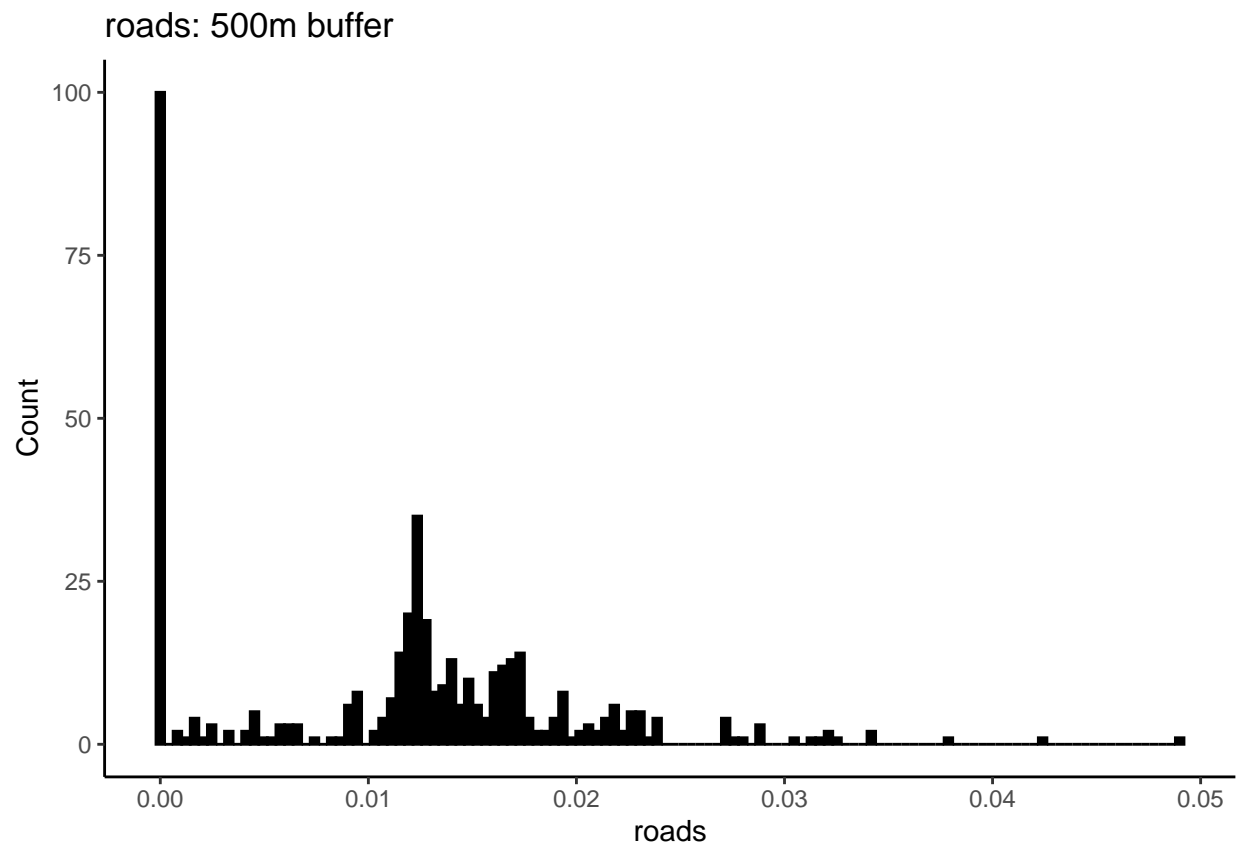


```
##  
## $pipe_trans
```

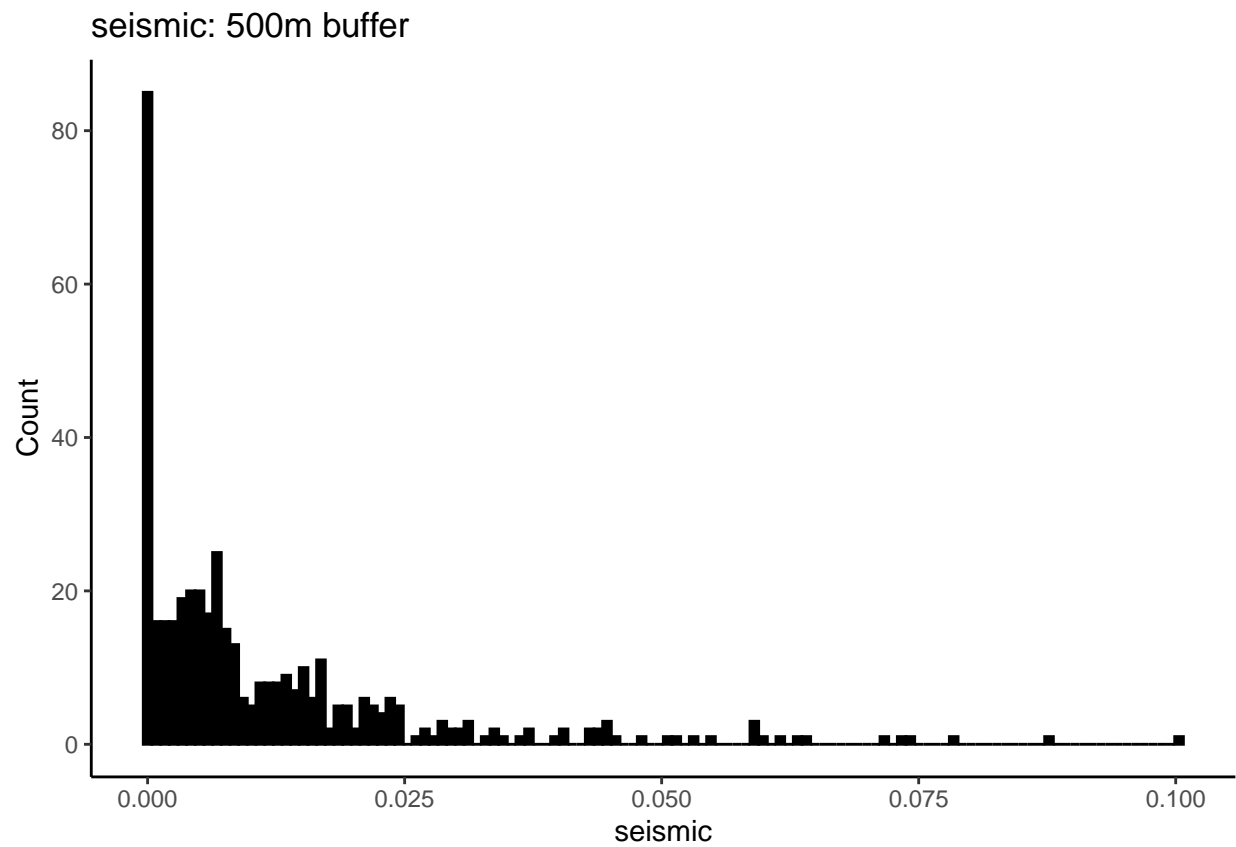


```
##  
## $roads
```

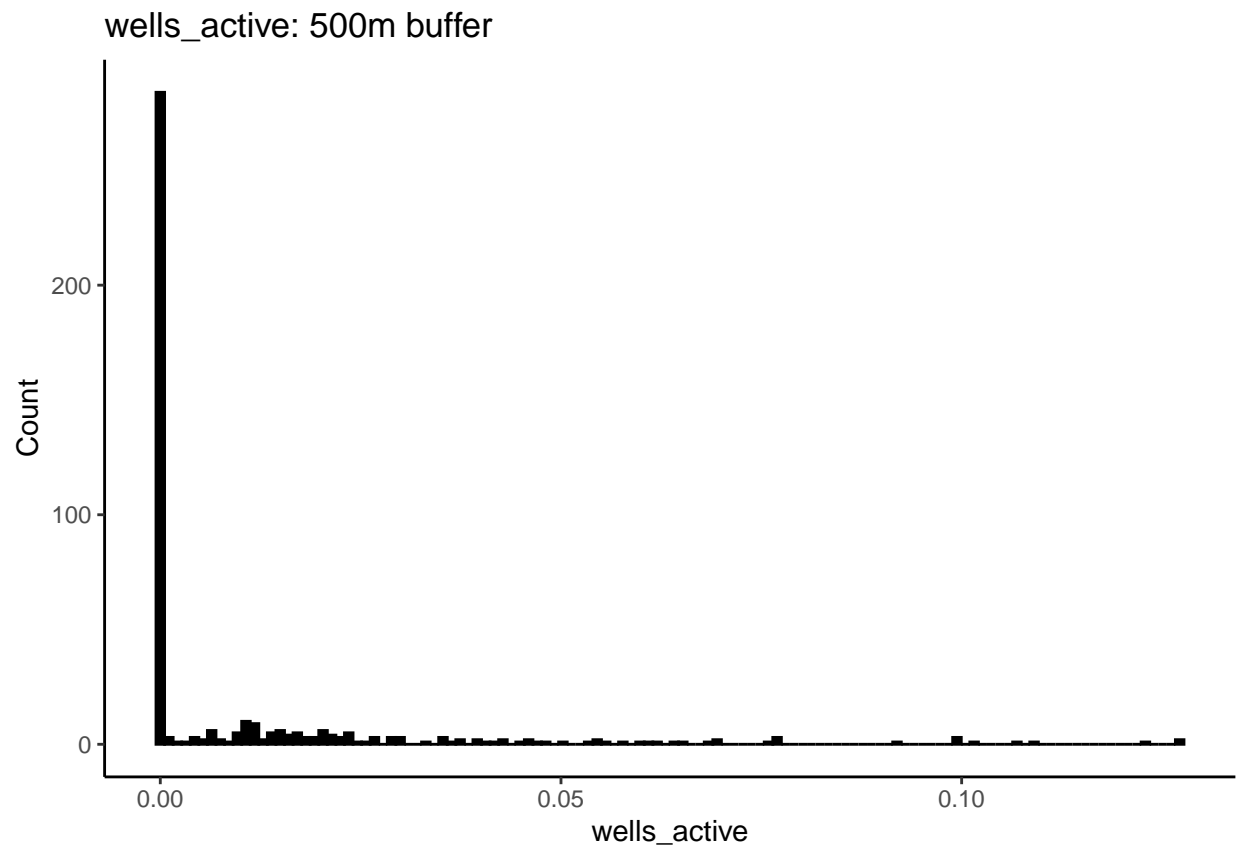




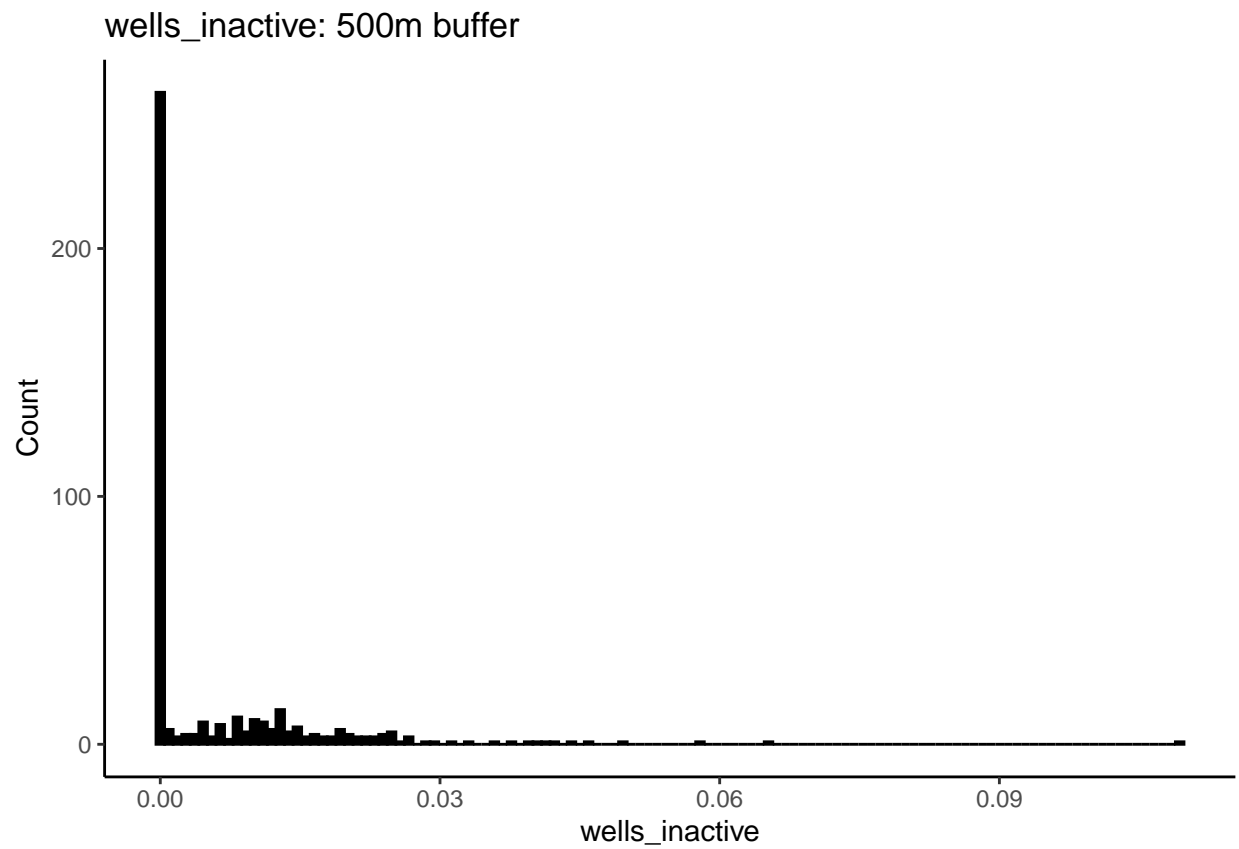
```
##  
## $seismic
```



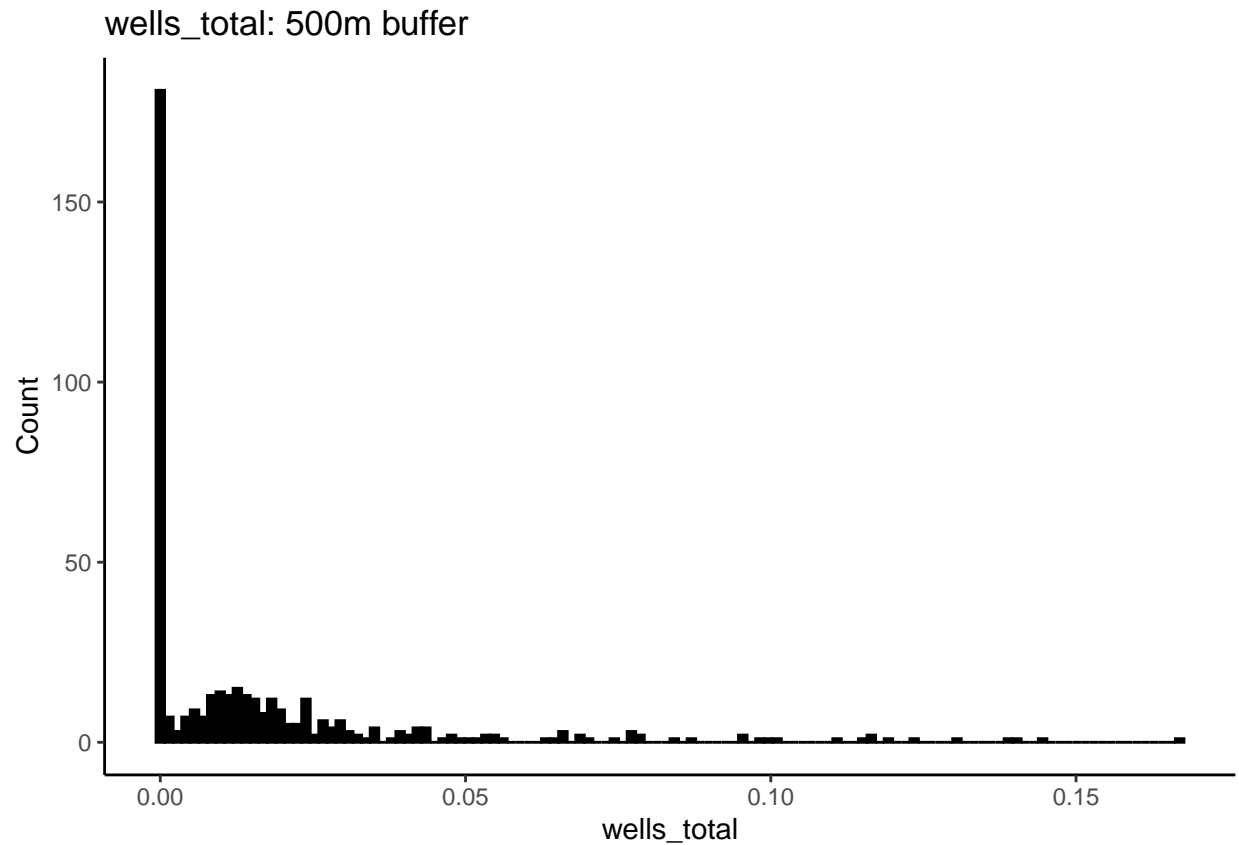
```
##  
## $wells_active
```



```
##  
## $wells_inactive
```

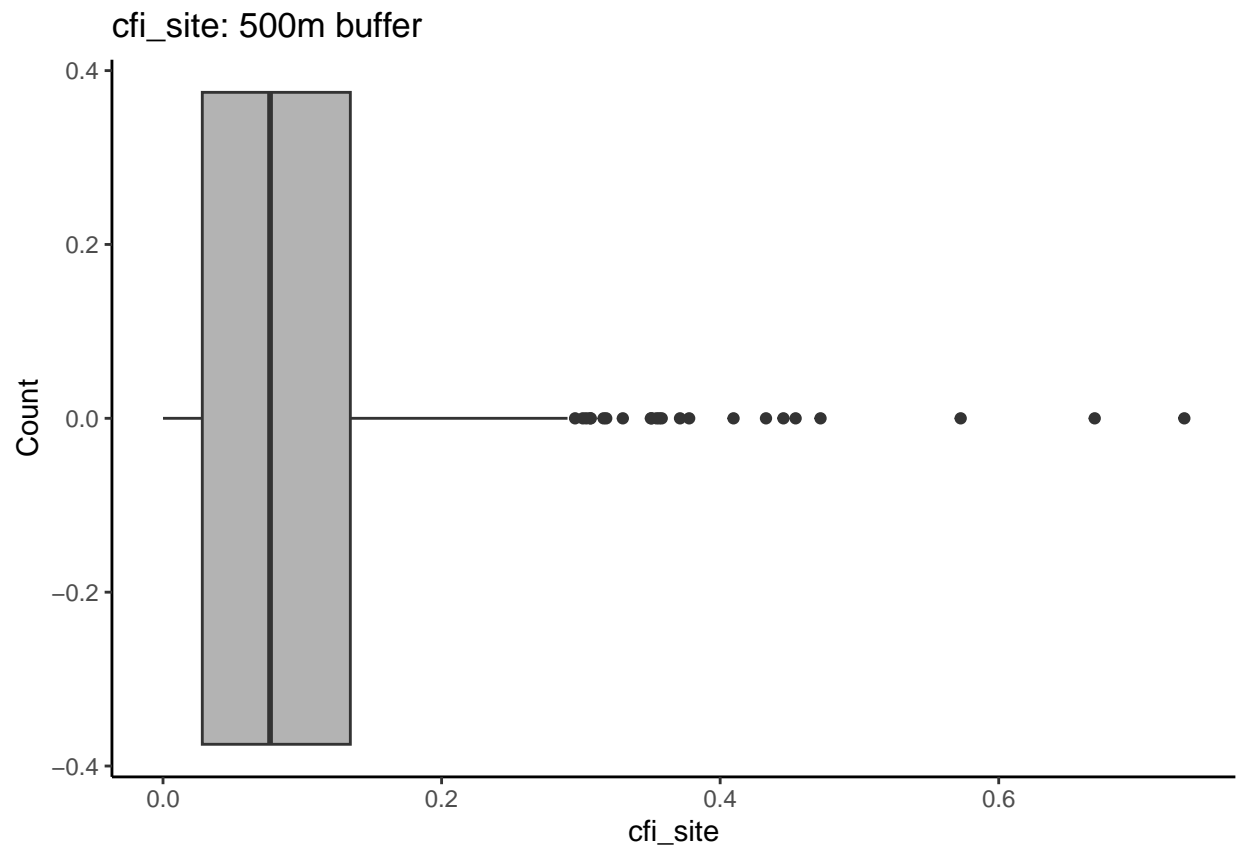


```
##  
## $wells_total
```

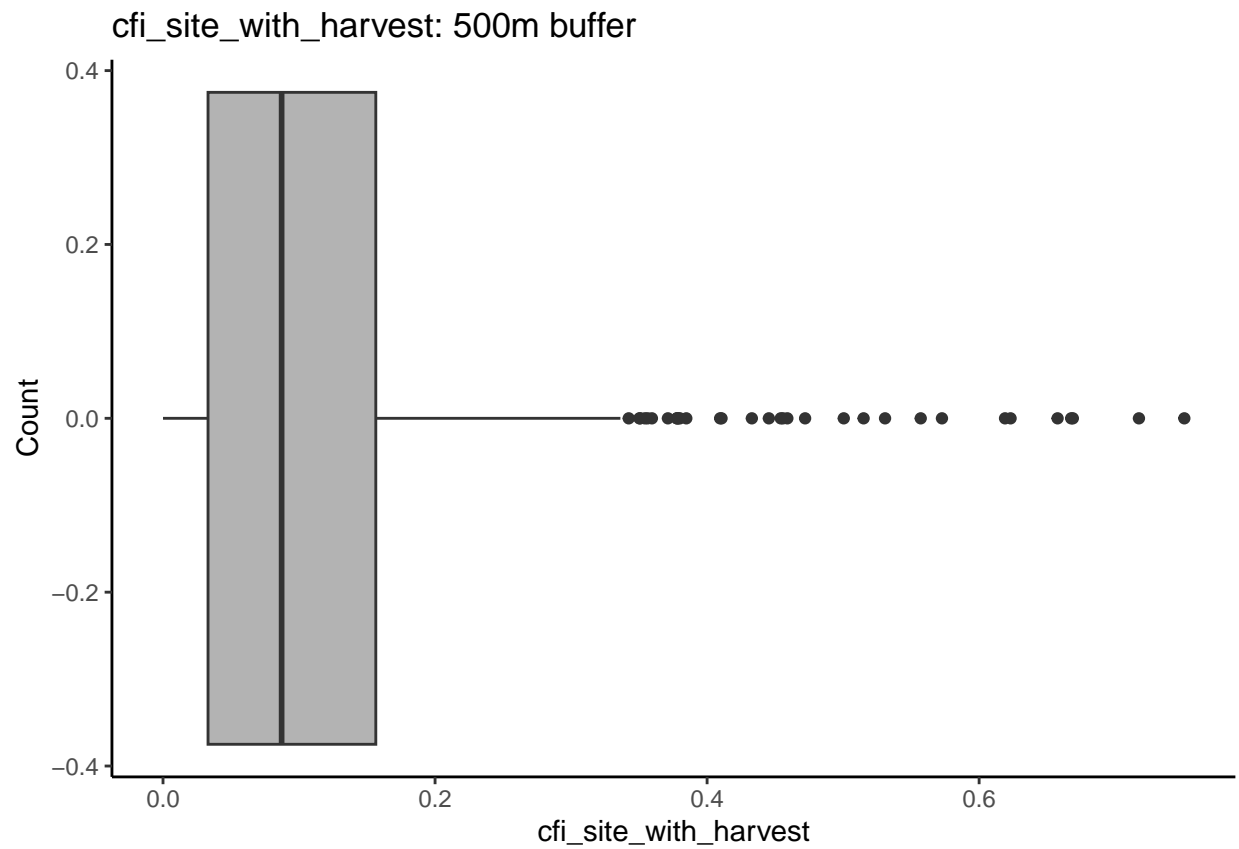


## 2.2. Box and whiskers at 500 meters:

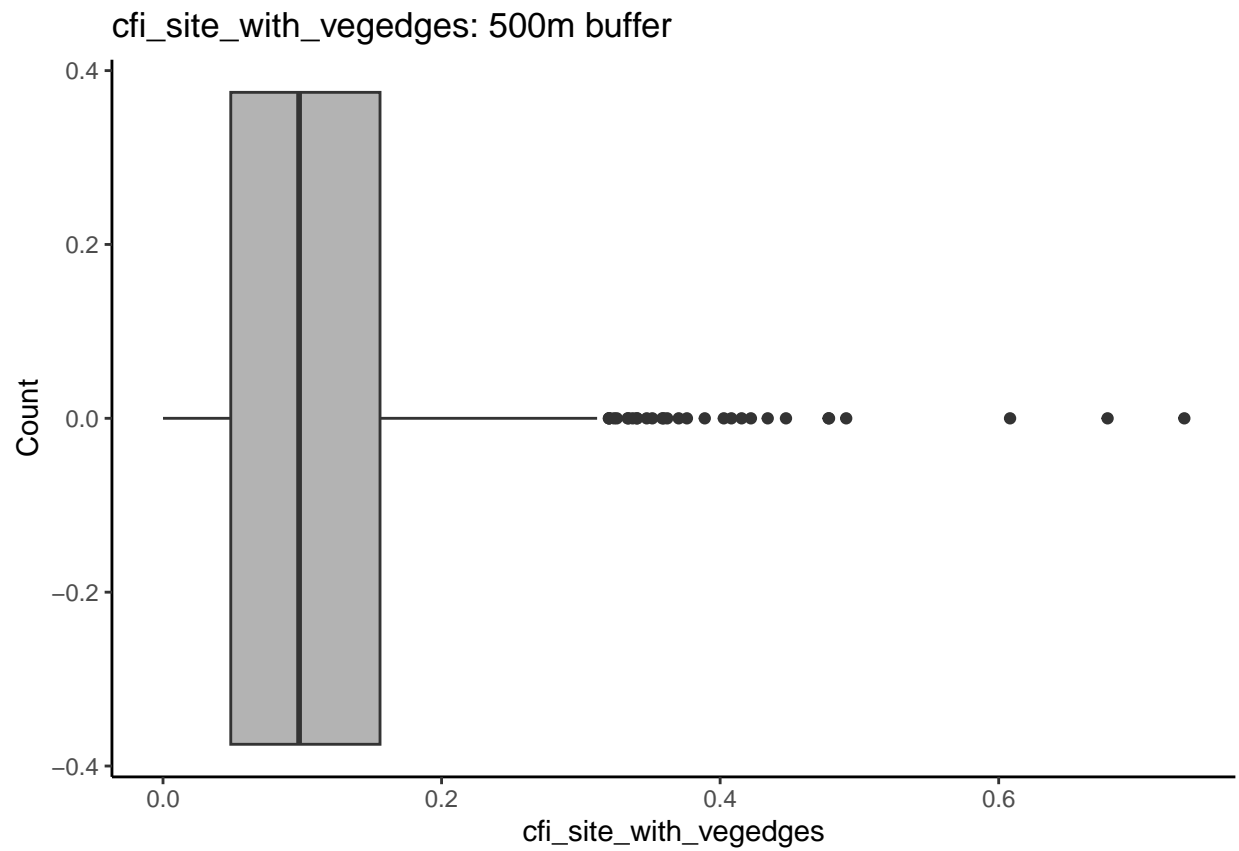
```
covs_to_explore %>%  
  
  filter(buffer_dist == 500) %>%  
  
  select(-(1:8)) %>%  
  
  # use imap which will retain both the data (x) and the variable names (y)  
  imap(~ ggplot(mapping = aes(x = .x)) +  
    geom_boxplot(fill = "grey70", color = "grey20") +  
    labs(title = paste0(.y, ": 500m buffer"), x = .y, y = "Count") +  
    theme_classic())  
  
## $cfi_site
```



```
##  
## $cfi_site_with_harvest
```

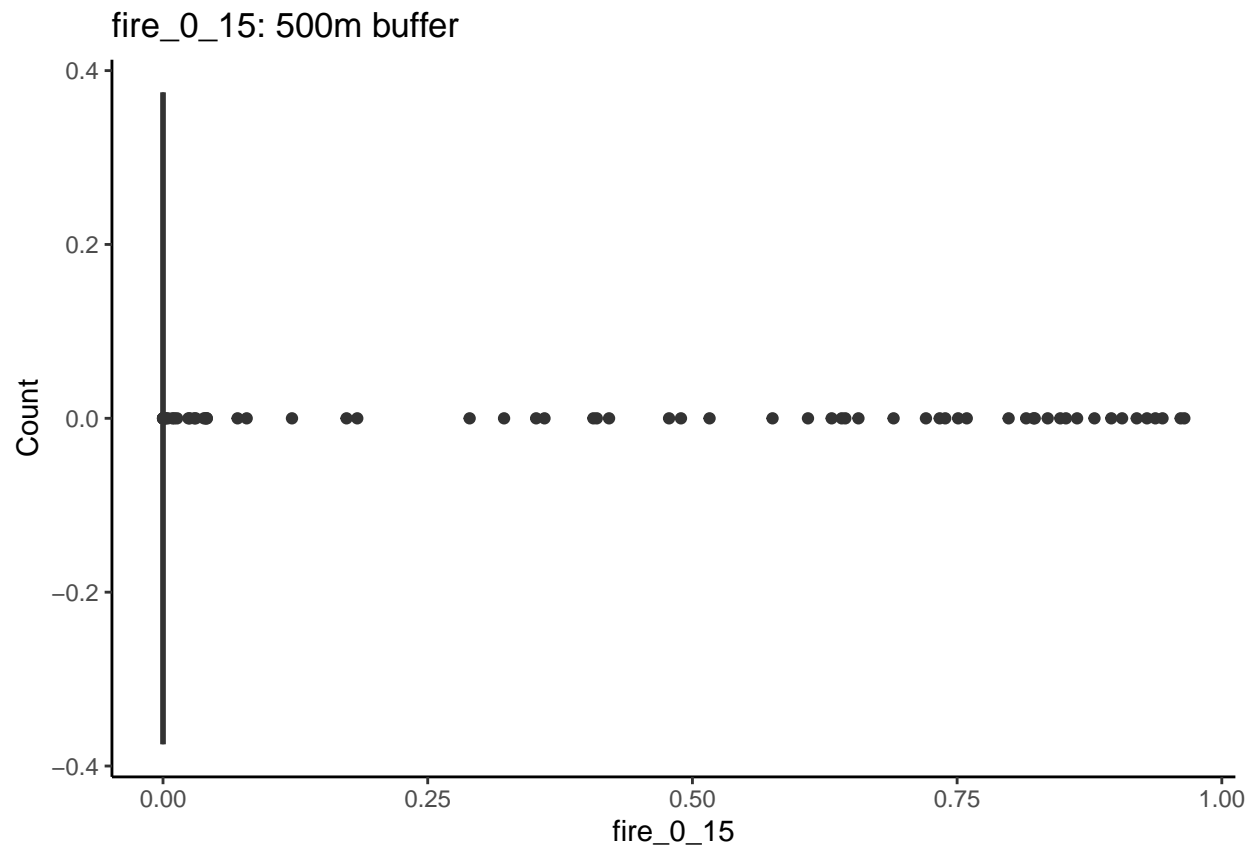


```
##  
## $cfi_site_with_vegedges
```

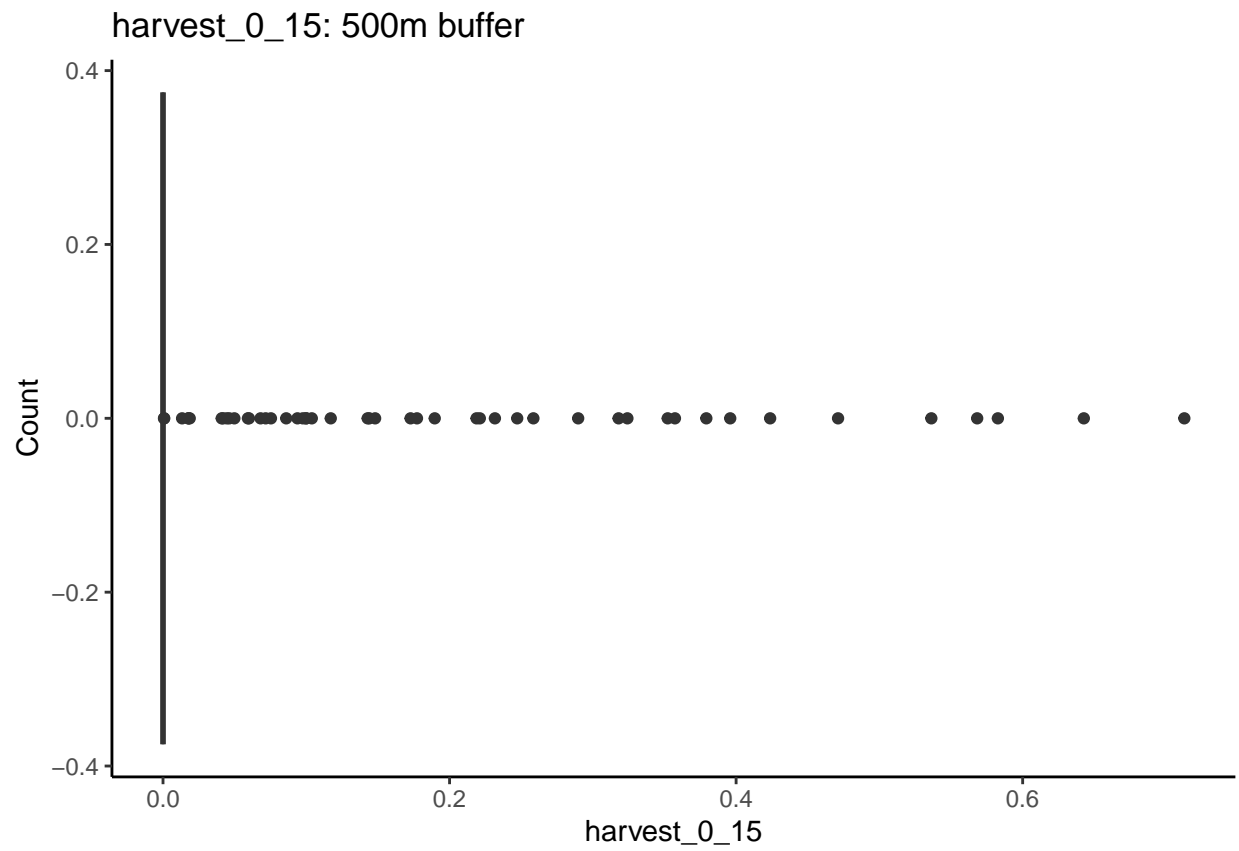


```
##  
## $fire_0_15
```

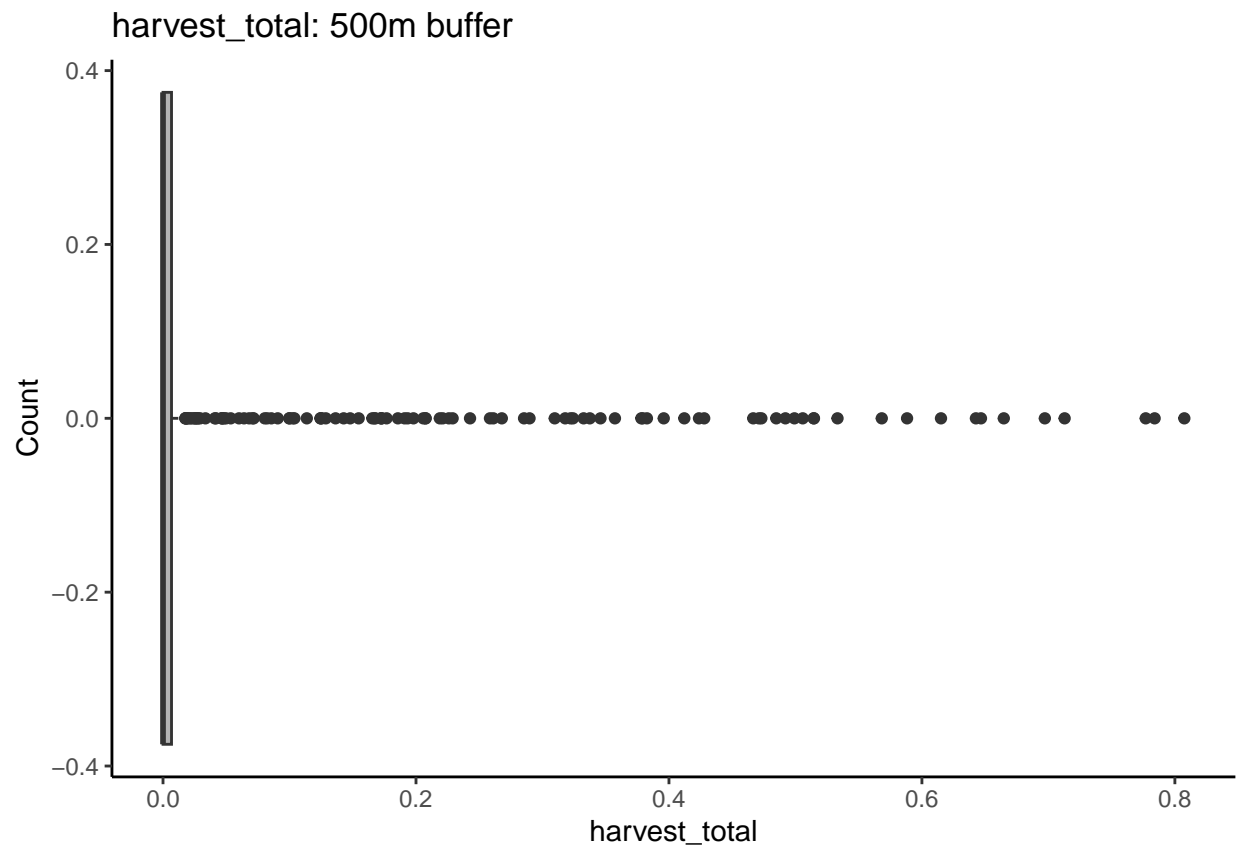




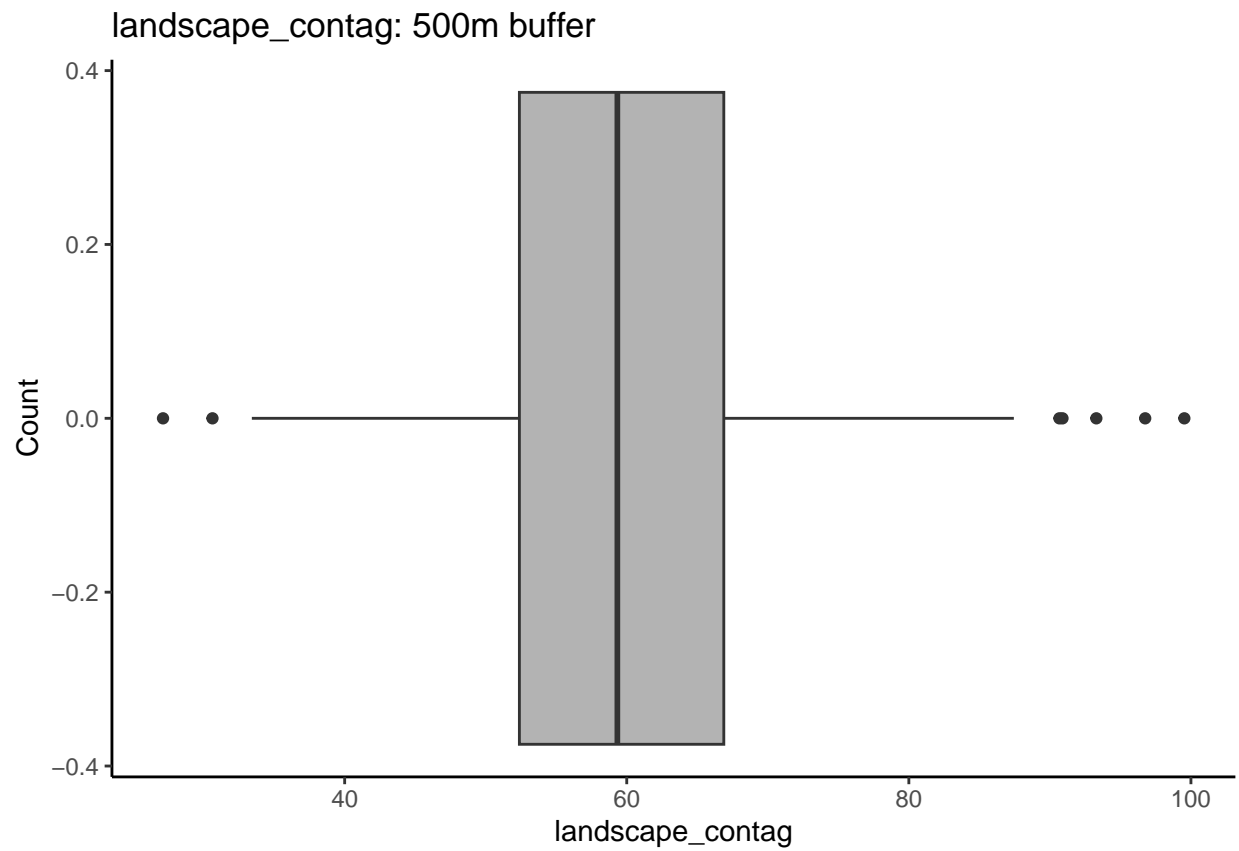
##  
## \$harvest\_0\_15



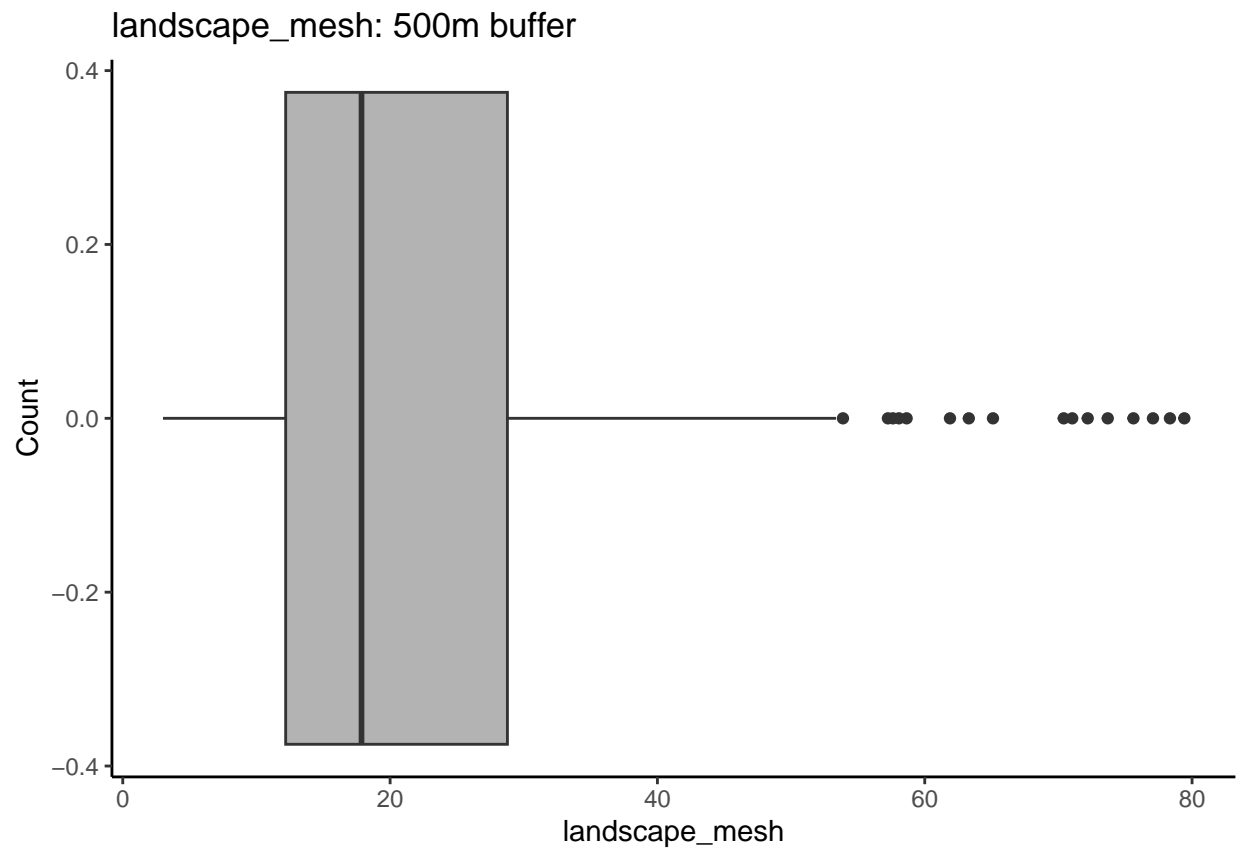
```
##  
## $harvest_total
```



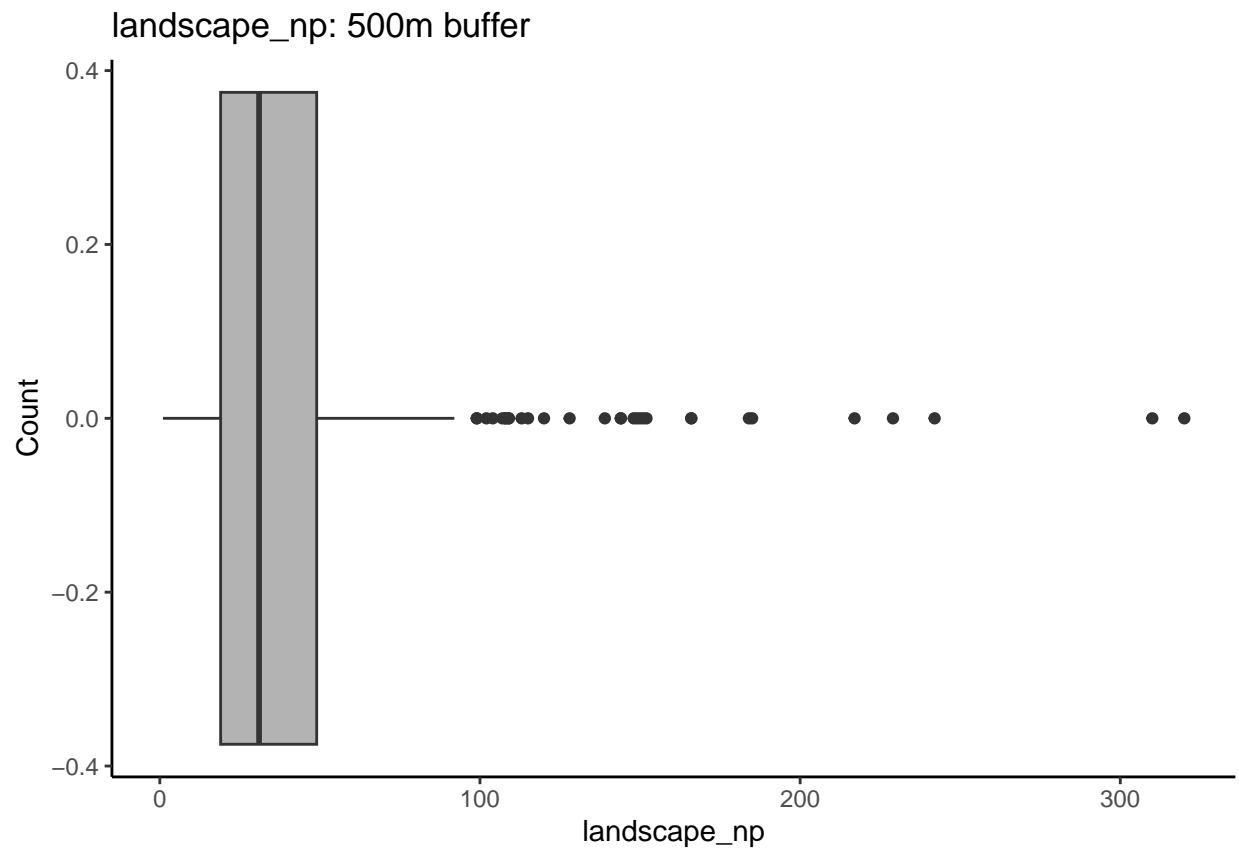
```
##  
## $landscape_contag
```



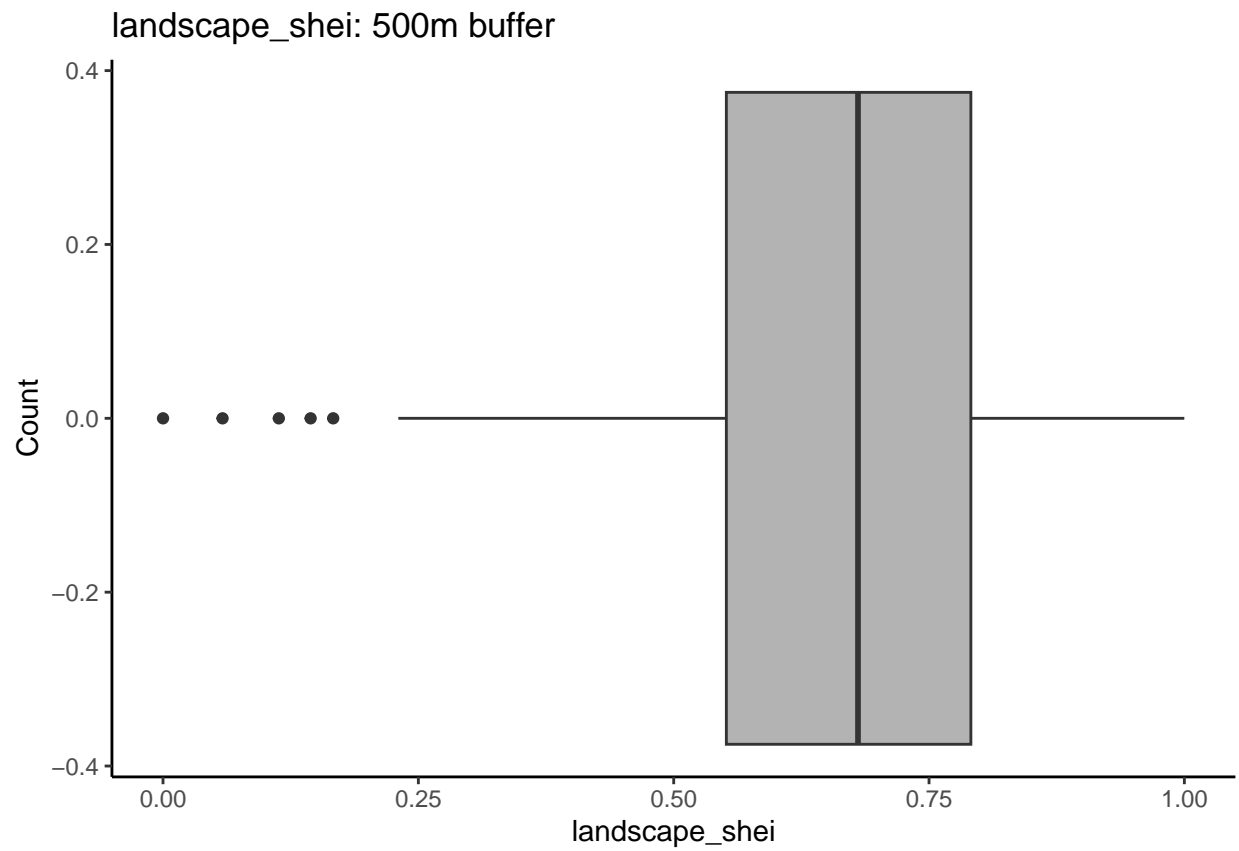
```
##  
## $landscape_mesh
```



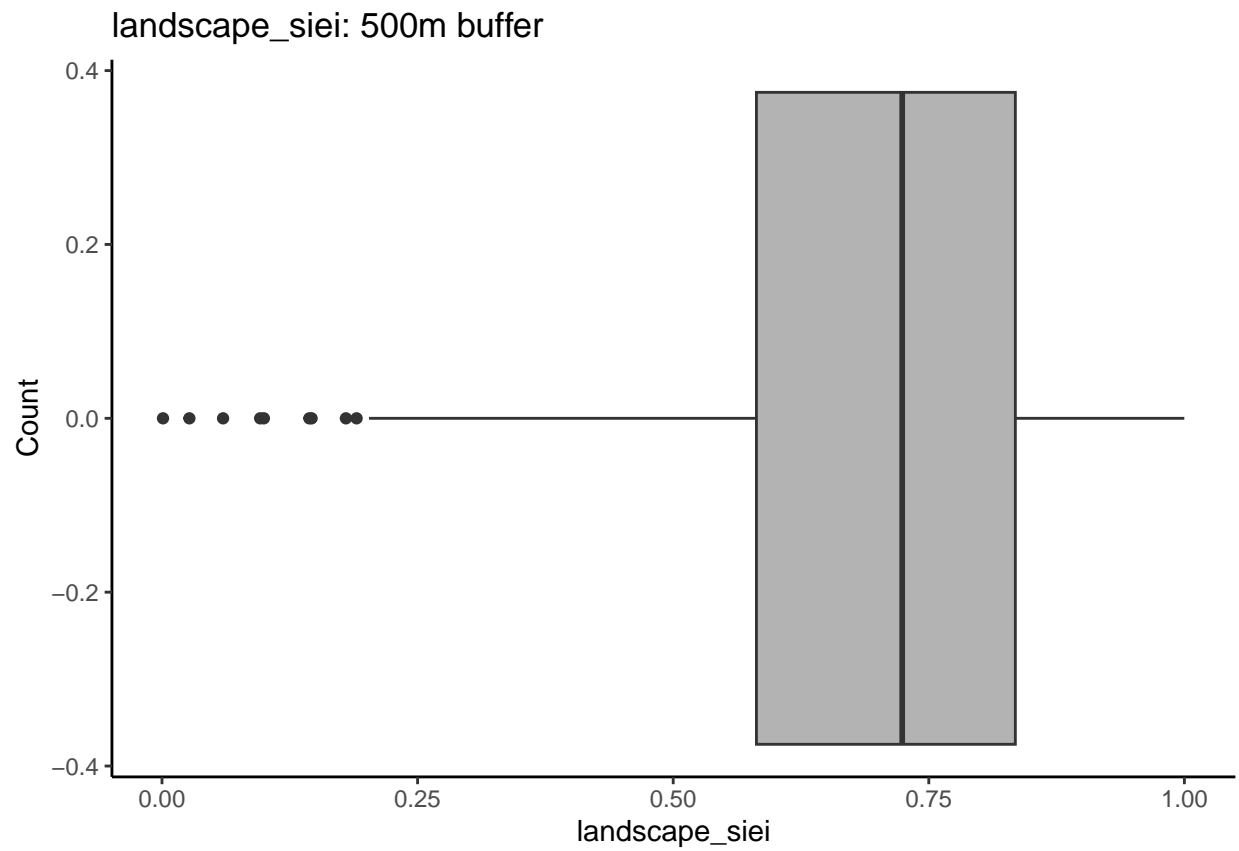
```
##  
## $landscape_np
```



```
##  
## $landscape_shei
```

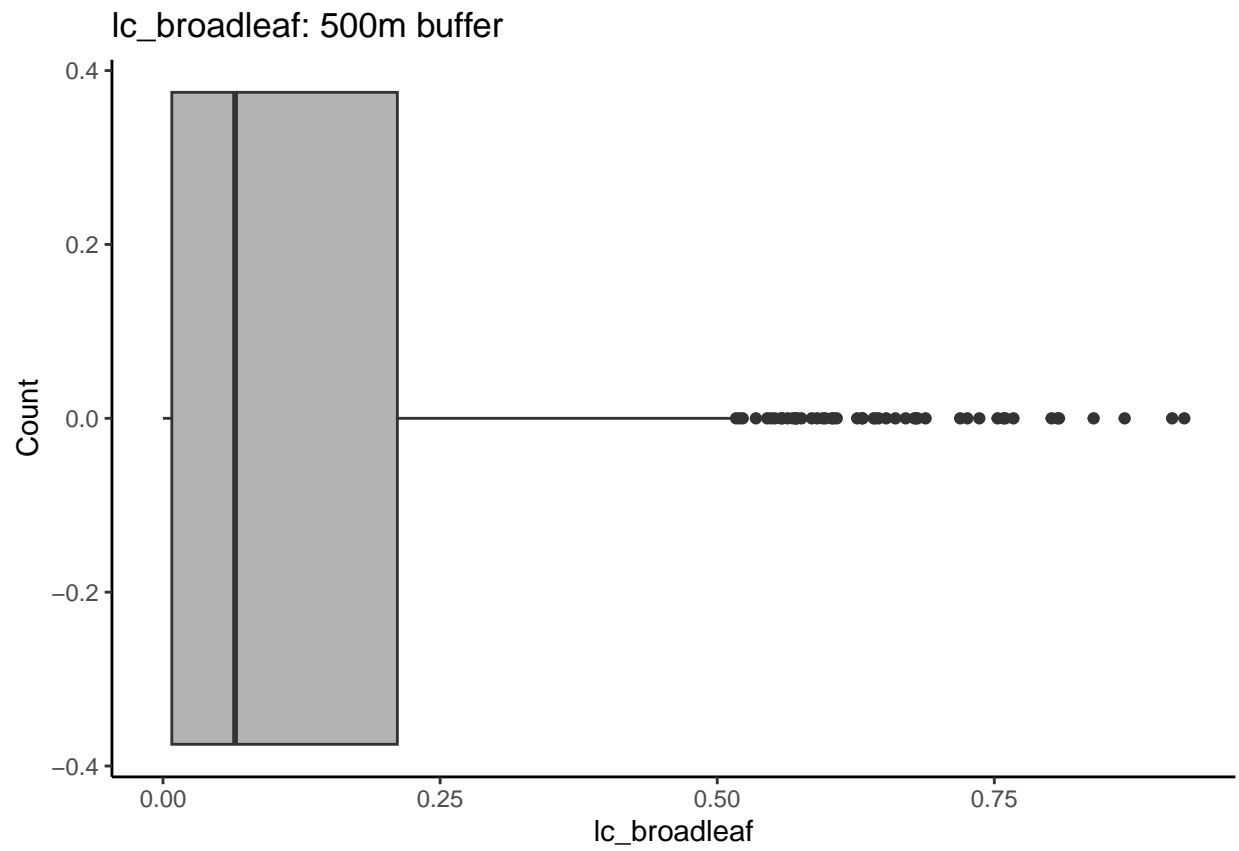


```
##  
## $landscape_siei
```

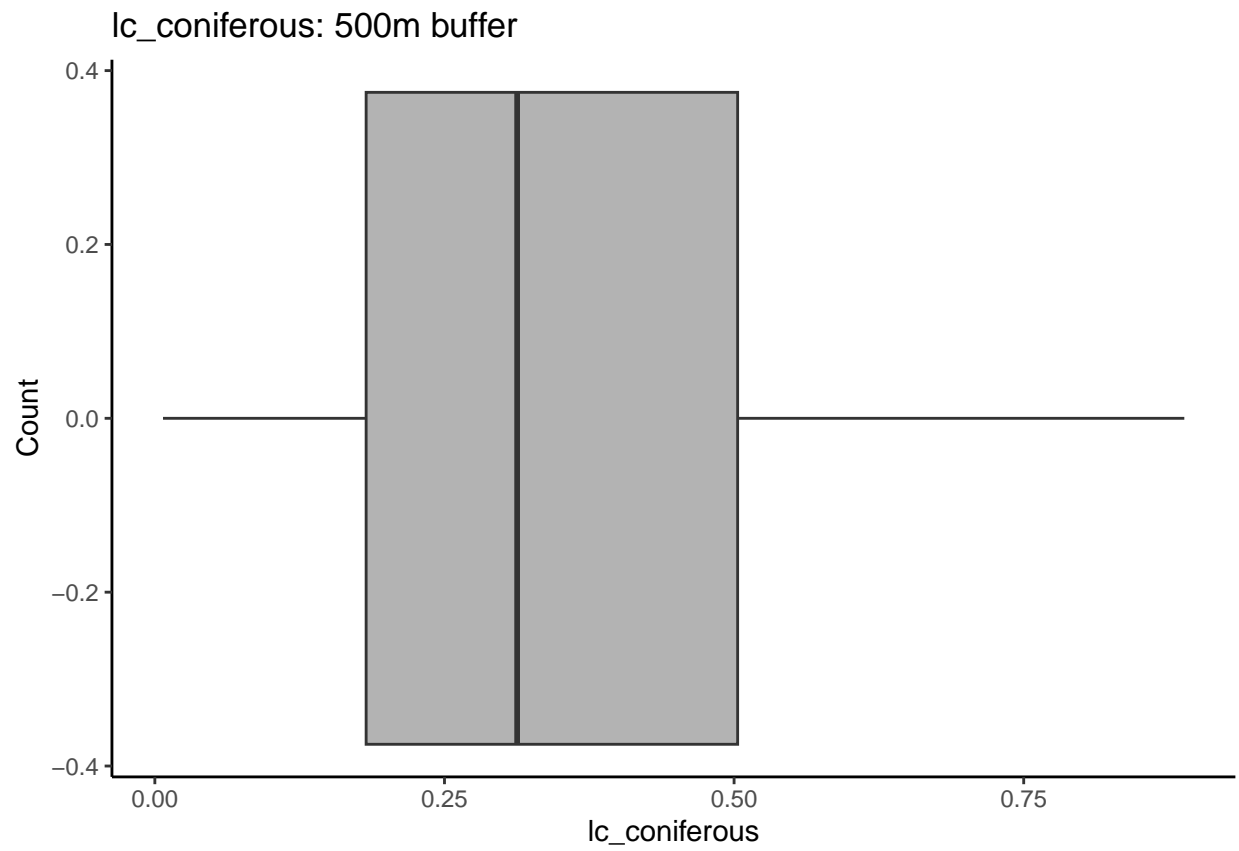


```
##  
## $lc_broadleaf
```

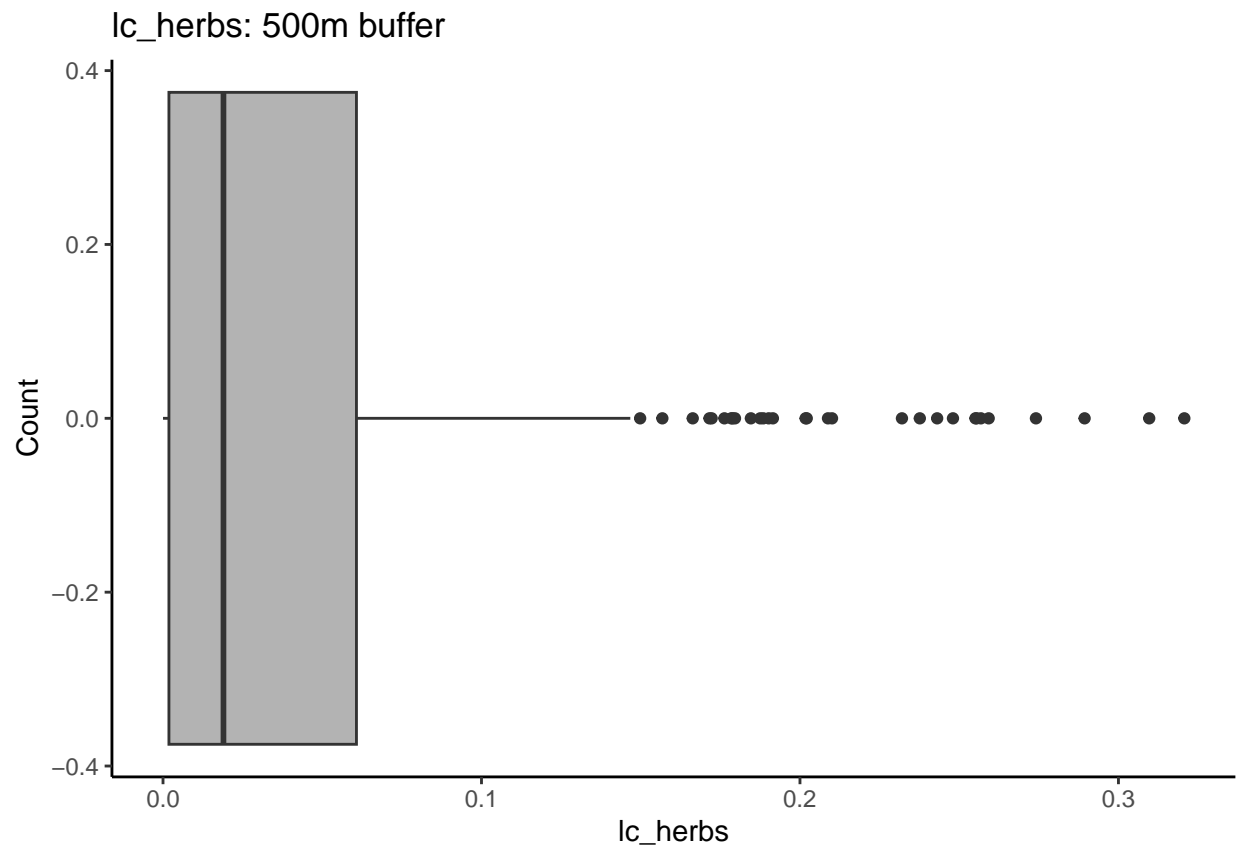




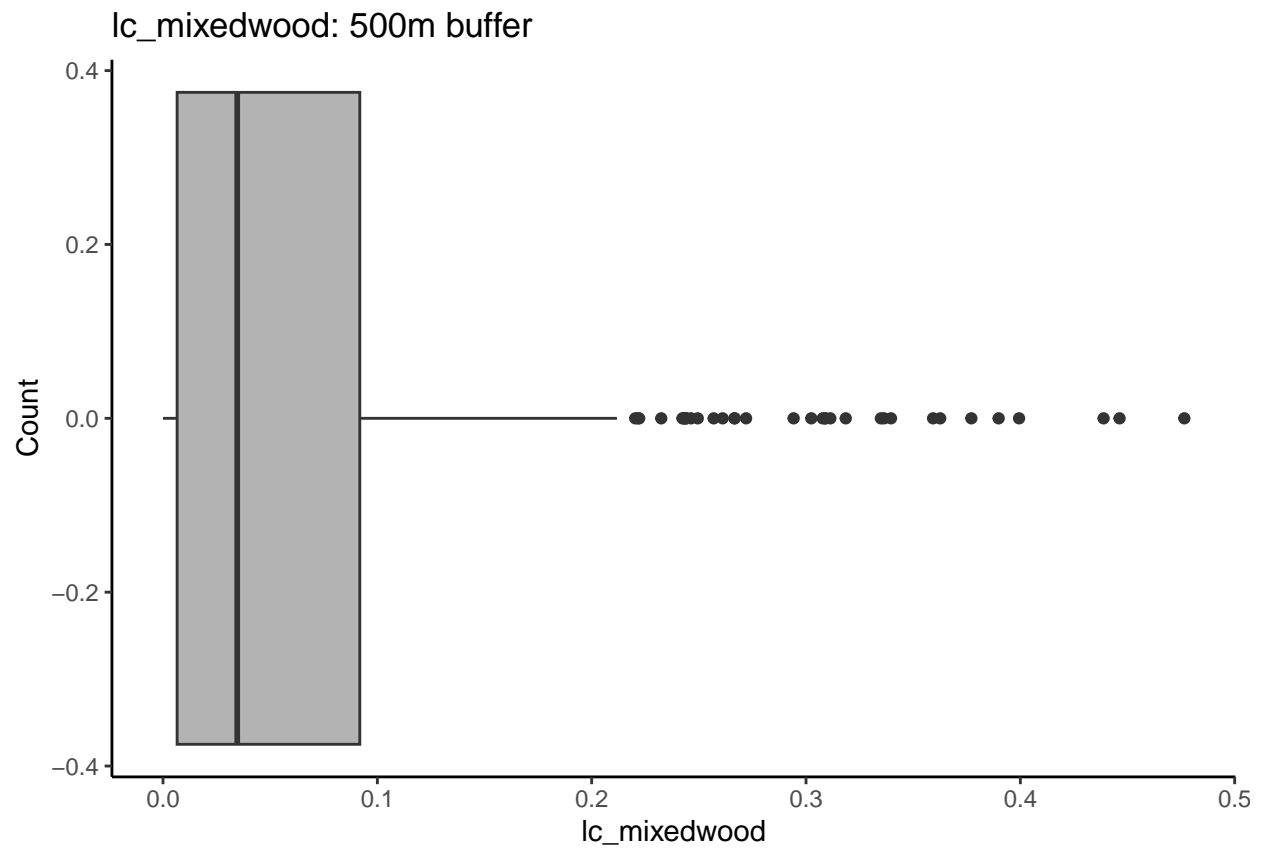
```
##  
## $lc_coniferous
```



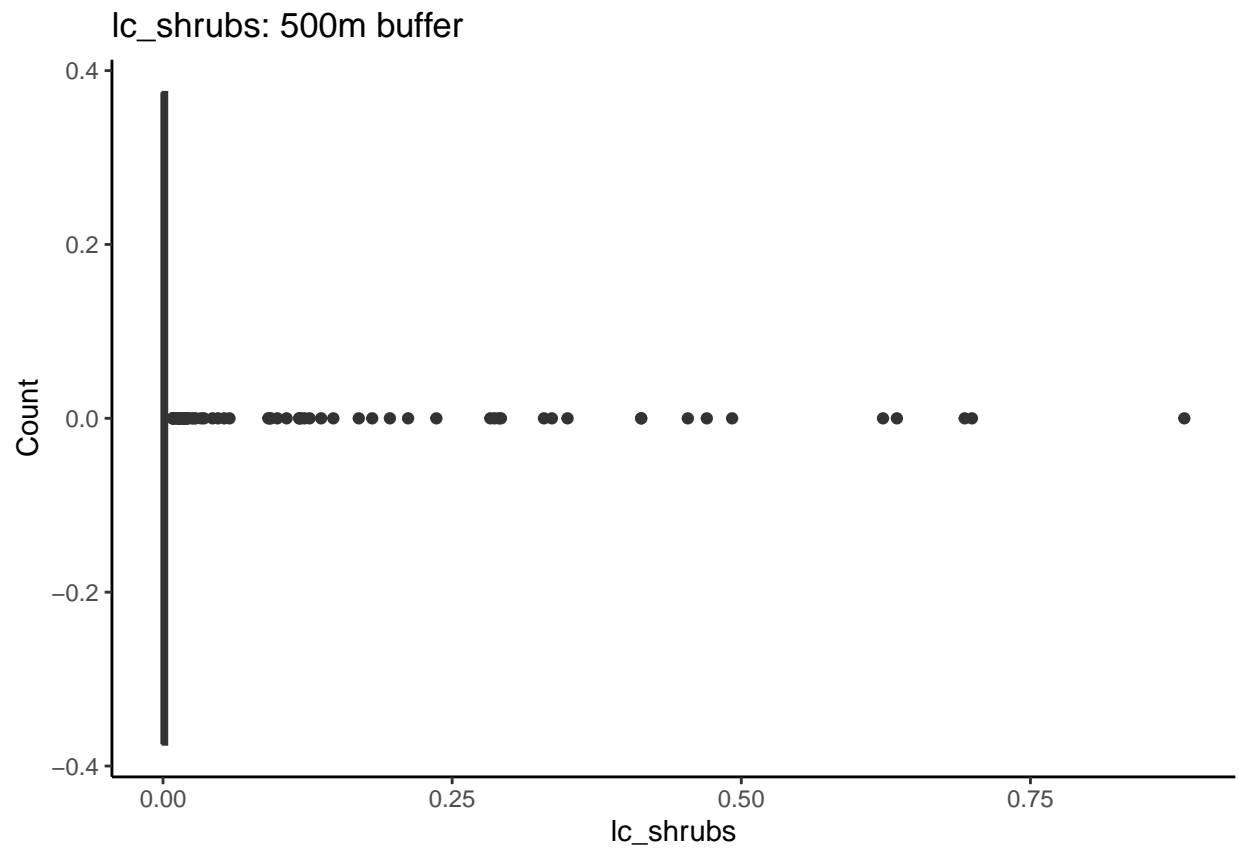
```
##  
## $lc_herbs
```



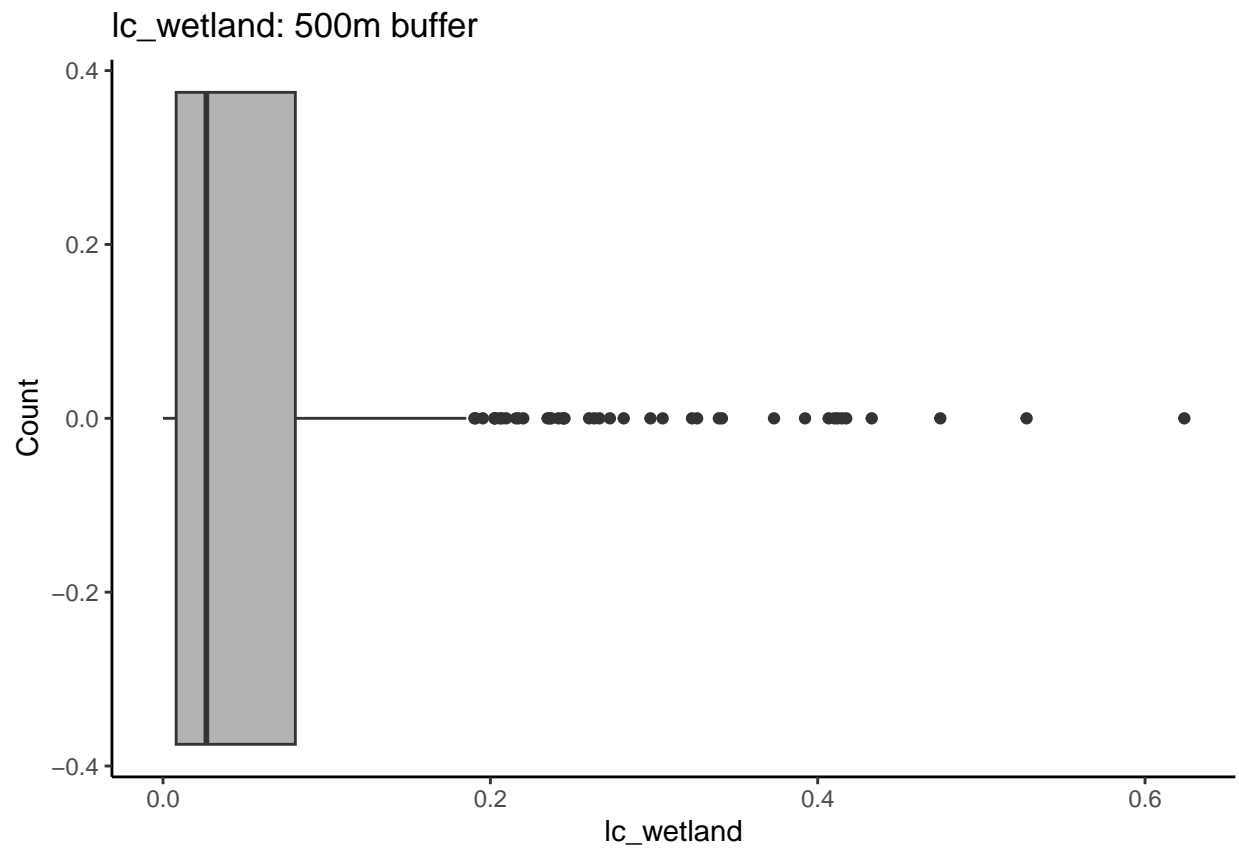
```
##  
## $lc_mixedwood
```



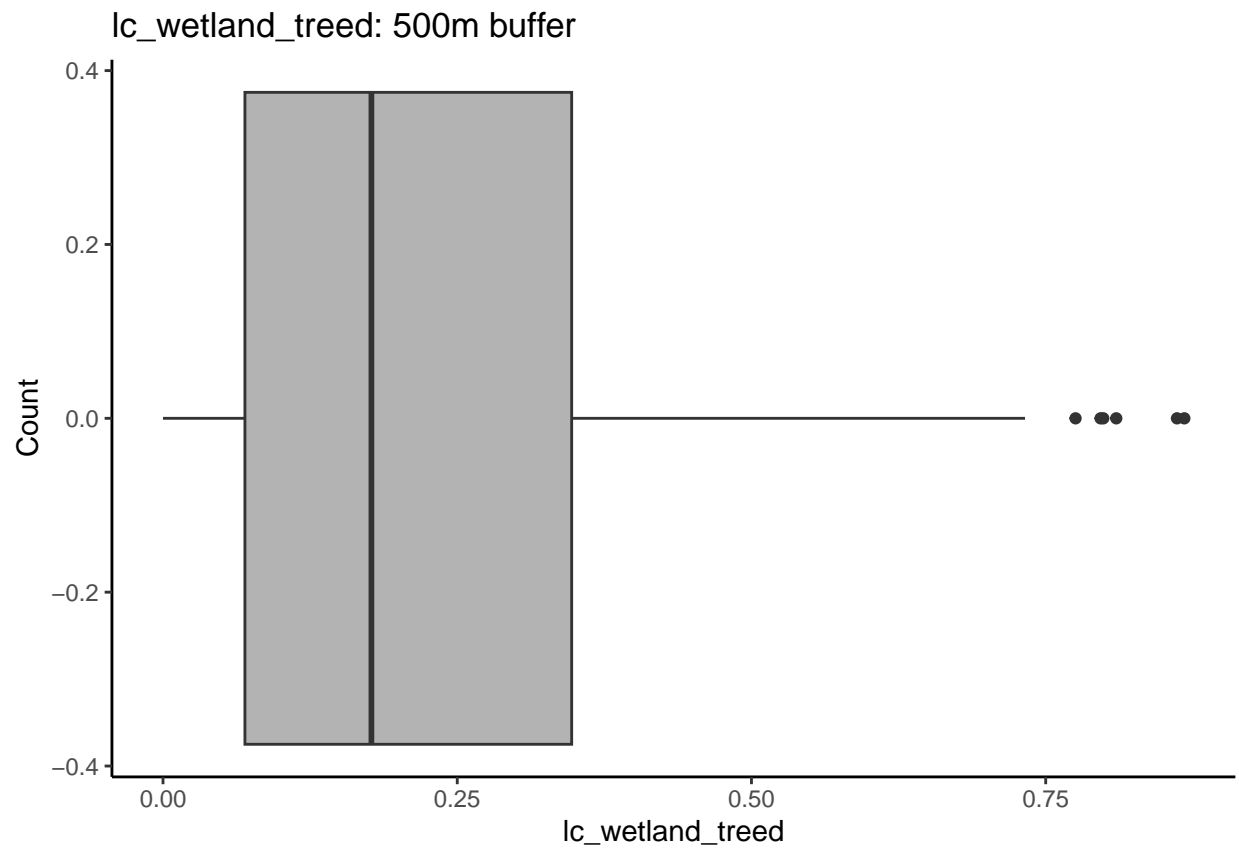
```
##  
## $lc_shrubs
```



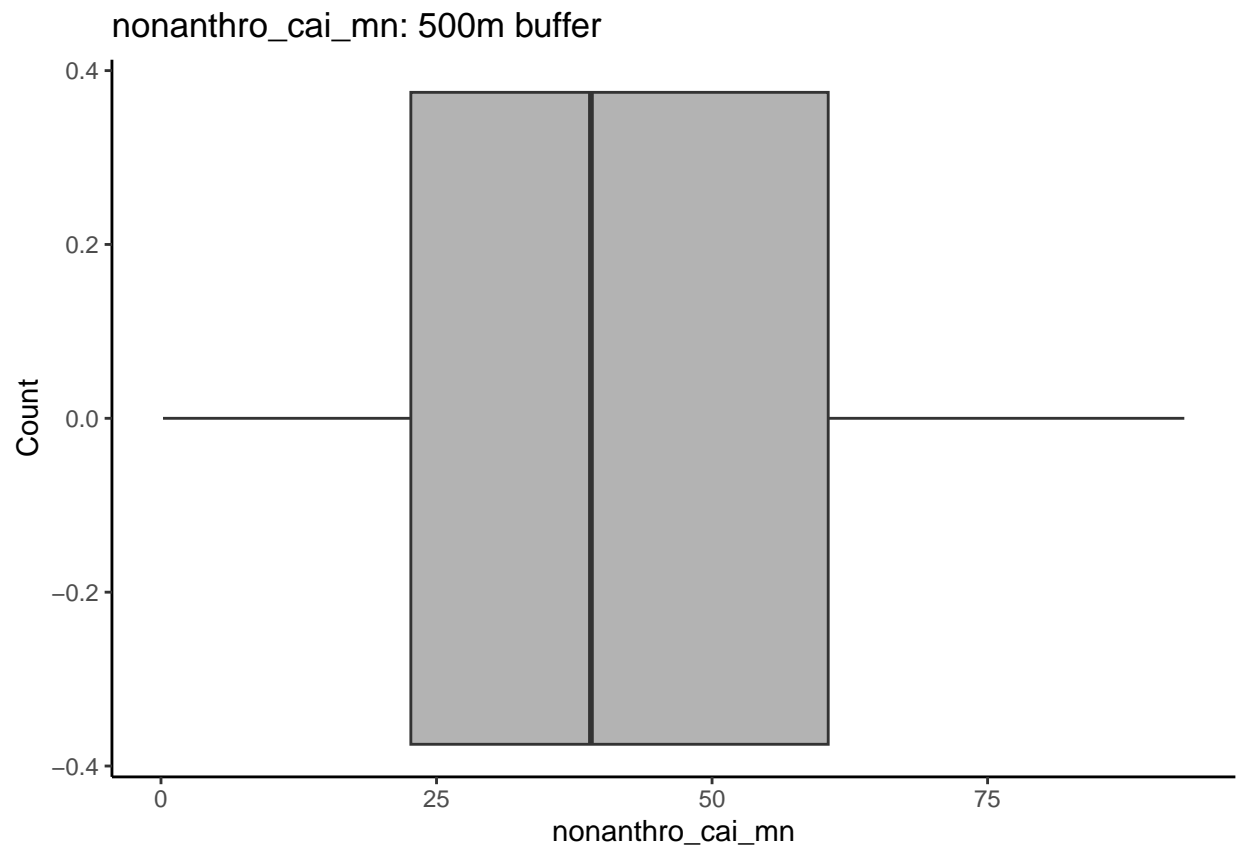
```
##  
## $lc_wetland
```



```
##  
## $lc_wetland_treed
```

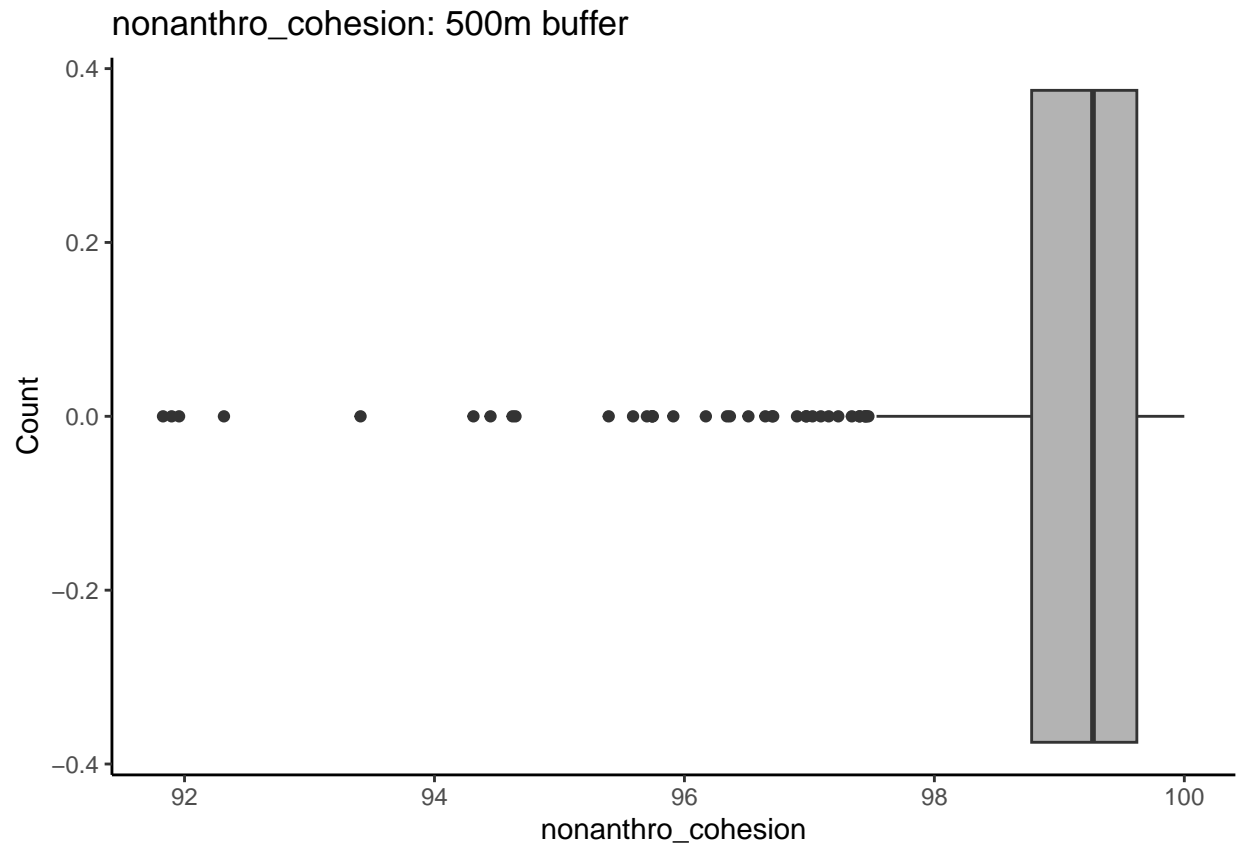


```
##  
## $nonanthro_cai_mn
```

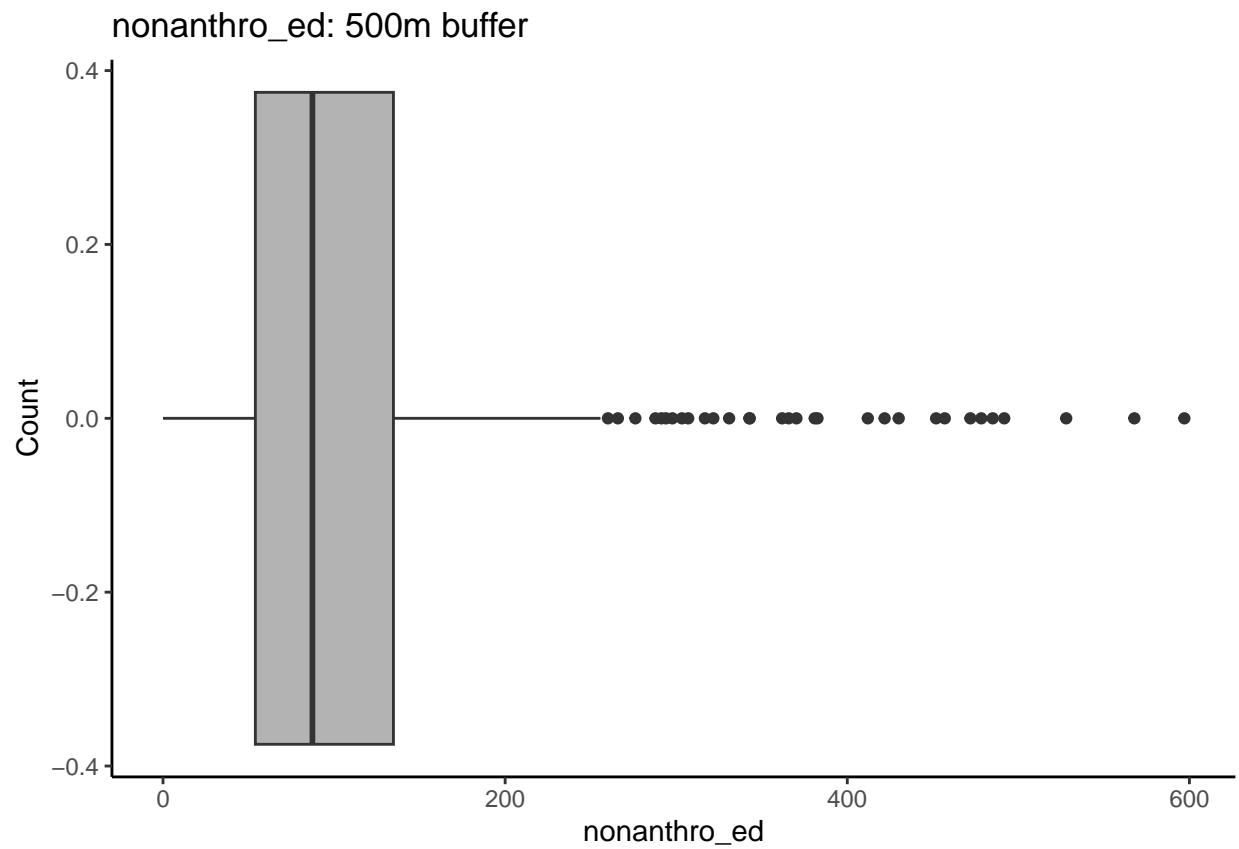


```
##  
## $nonanthro_cohesion
```

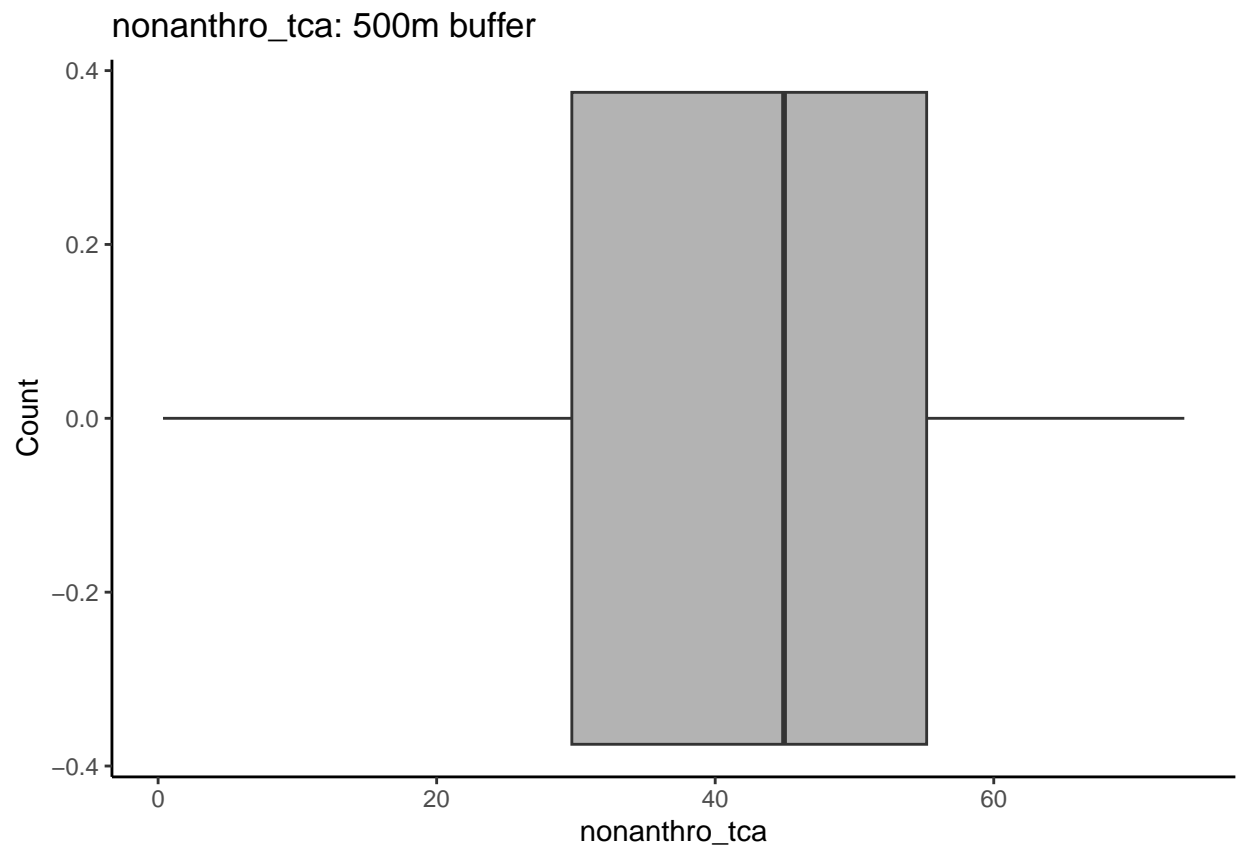




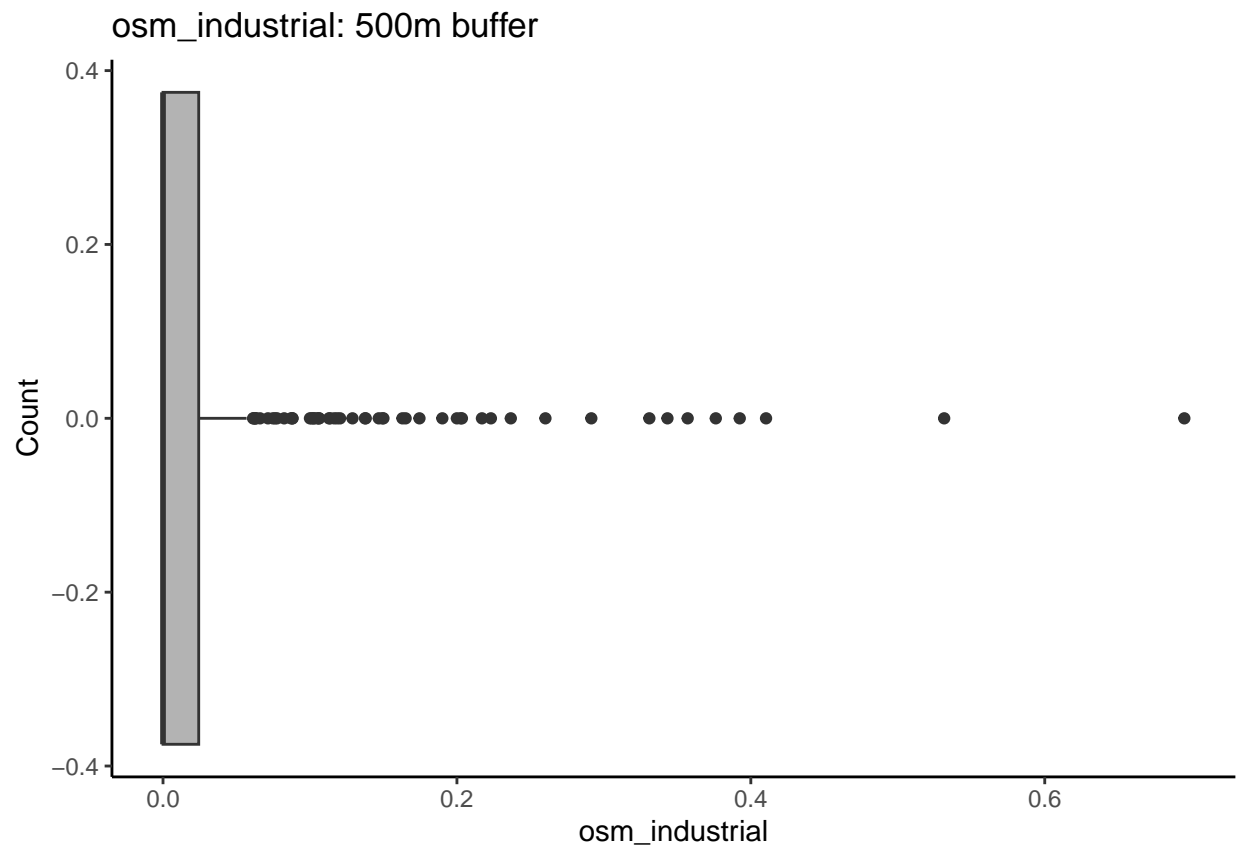
```
##  
## $nonanthro_ed
```



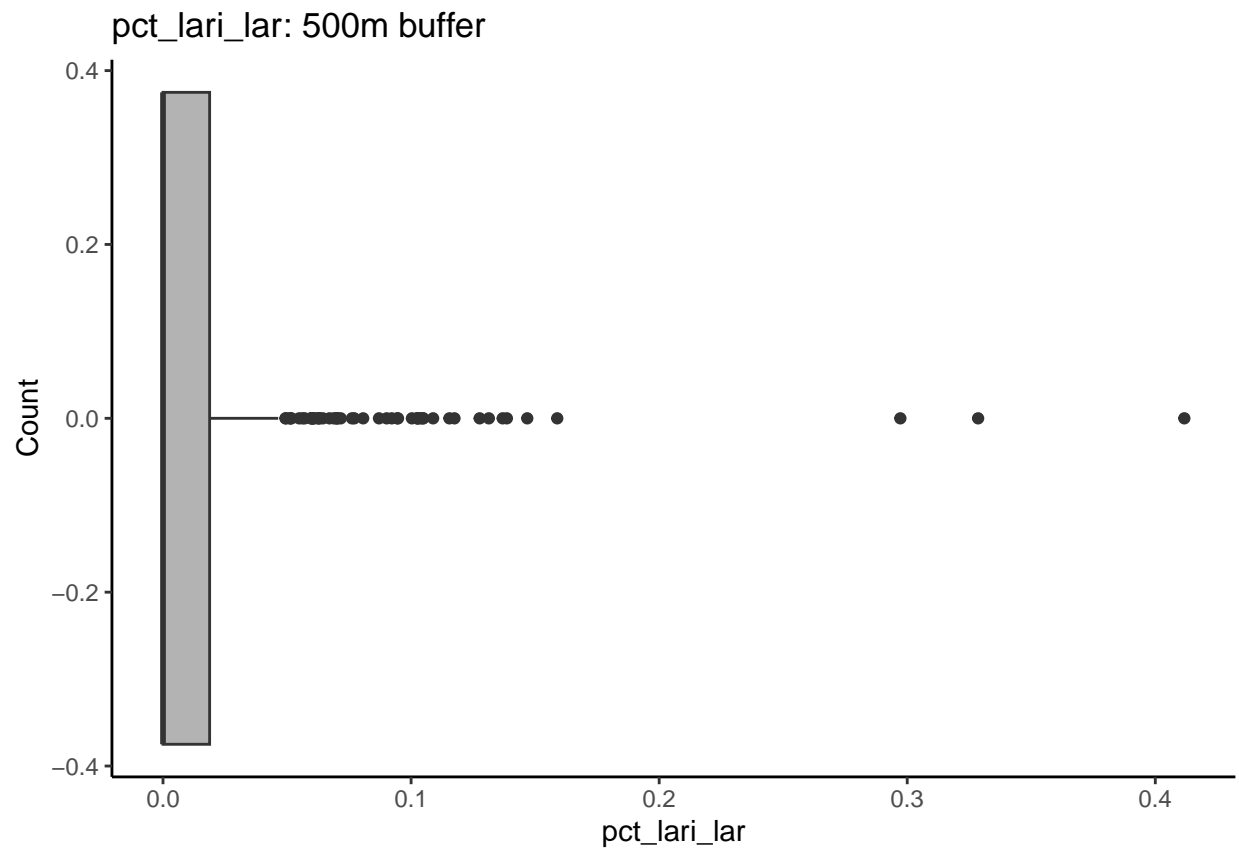
```
##  
## $nonanthro_tca
```



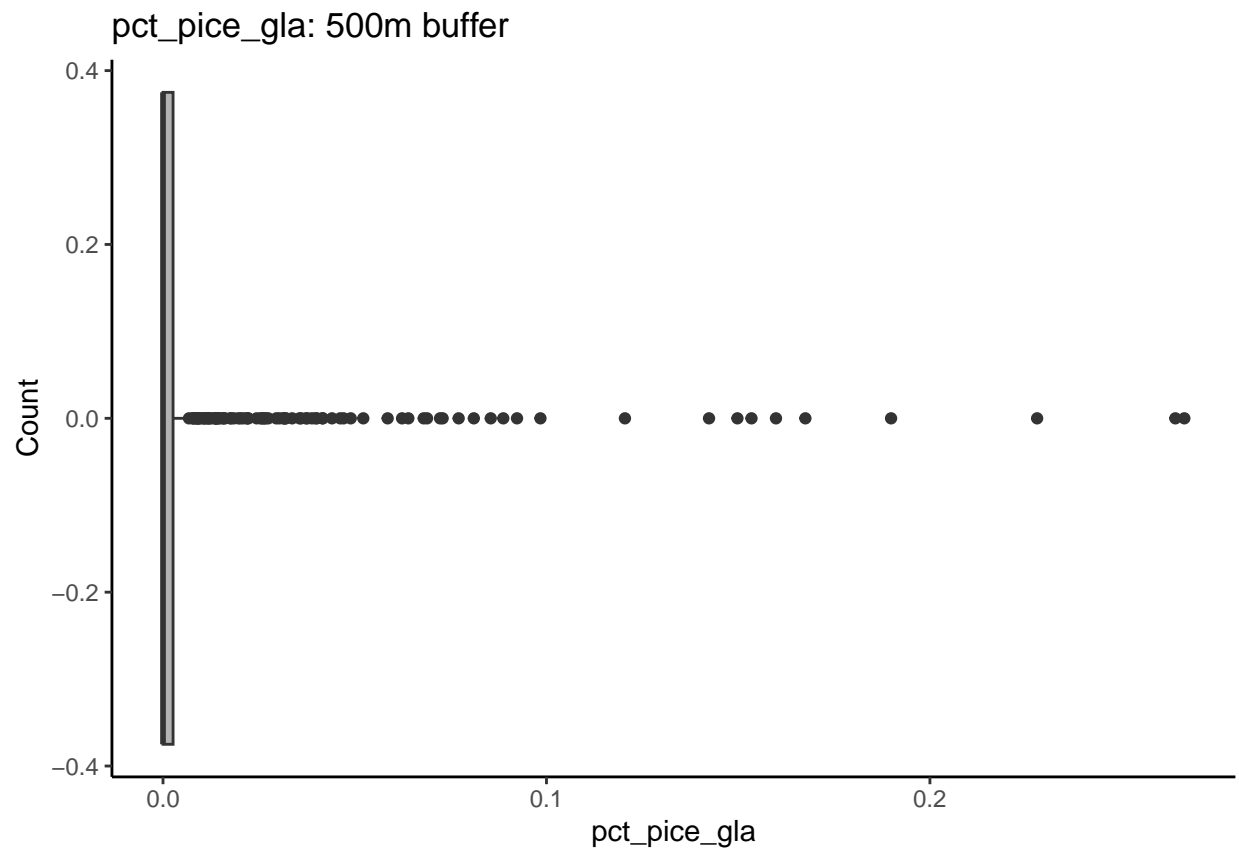
```
##  
## $osm_industrial
```



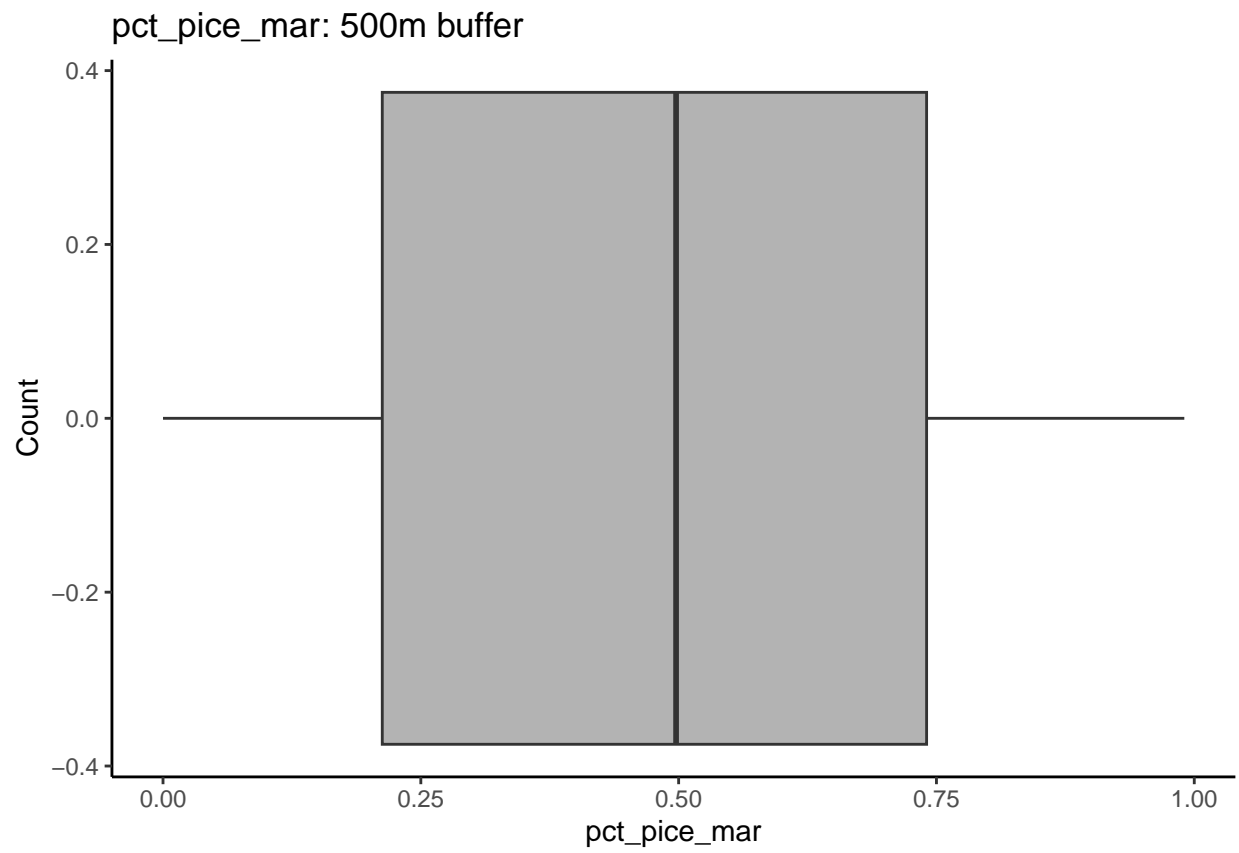
```
##  
## $pct_lari_lar
```



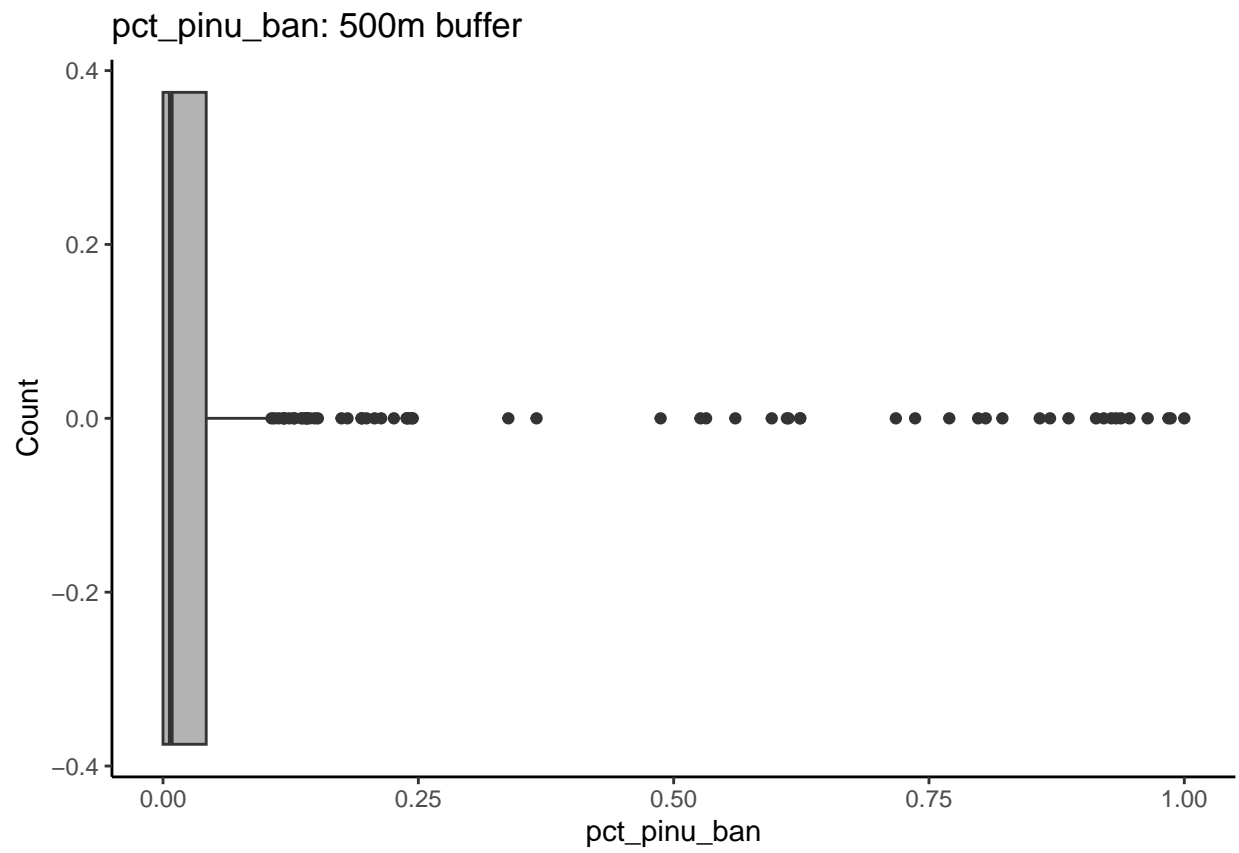
```
##  
## $pct_pice_gla
```



```
##  
## $pct_pice_mar
```

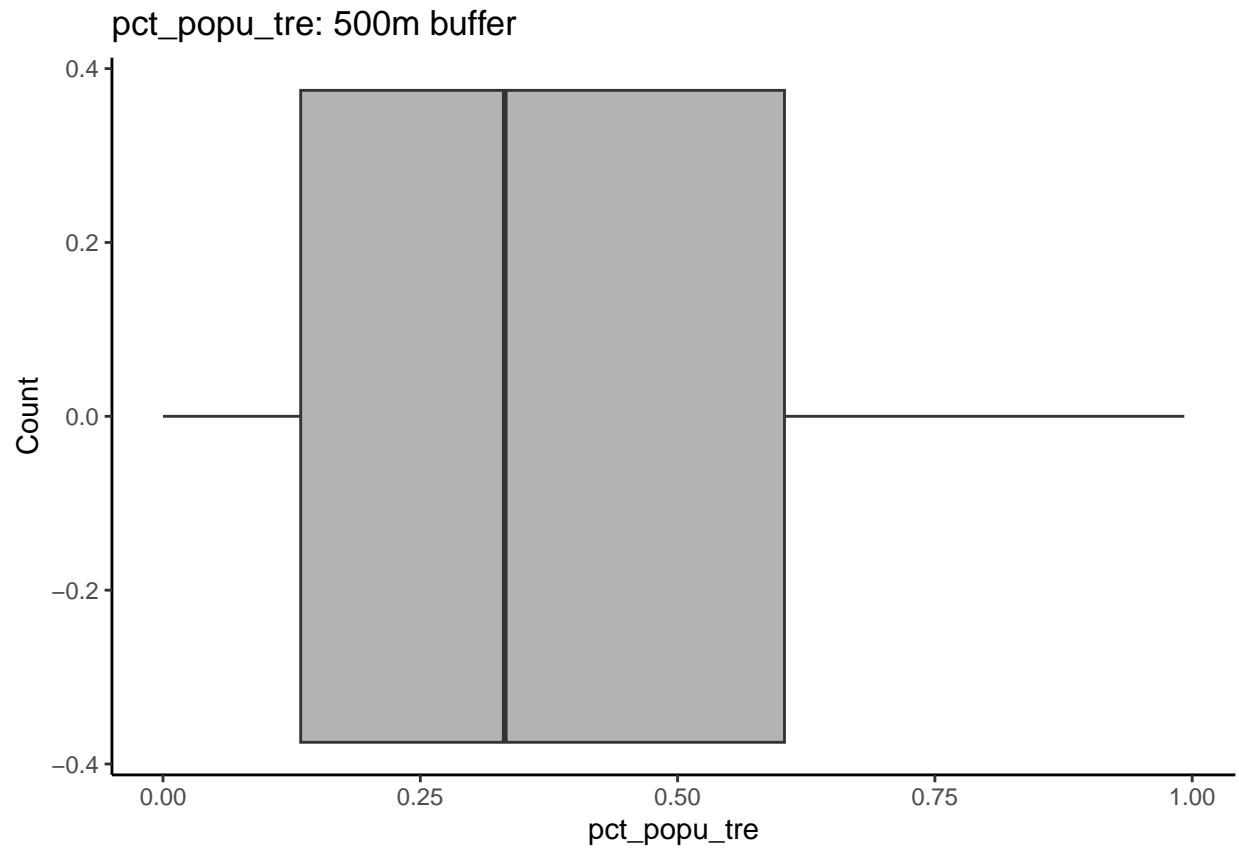


```
##  
## $pct_pinu_ban
```

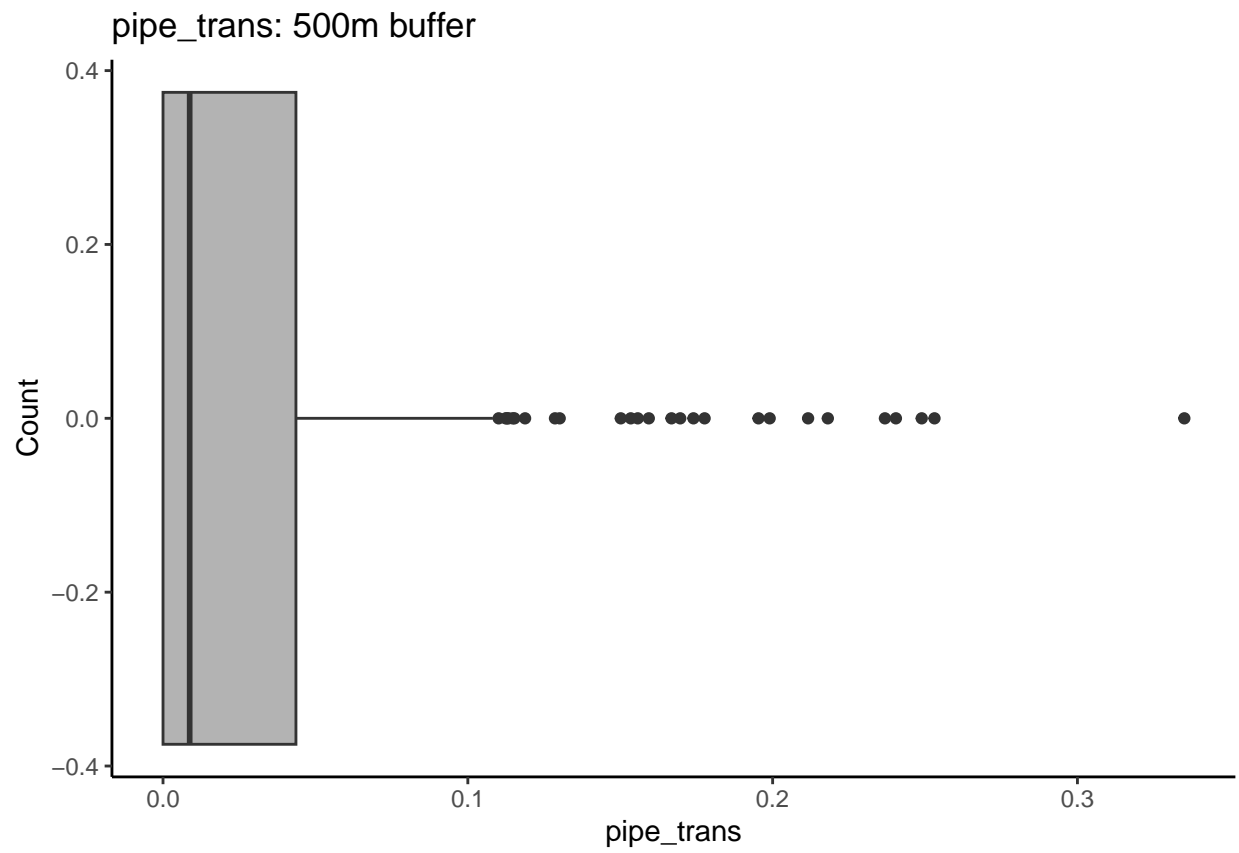


```
##  
## $pct_popu_tre
```

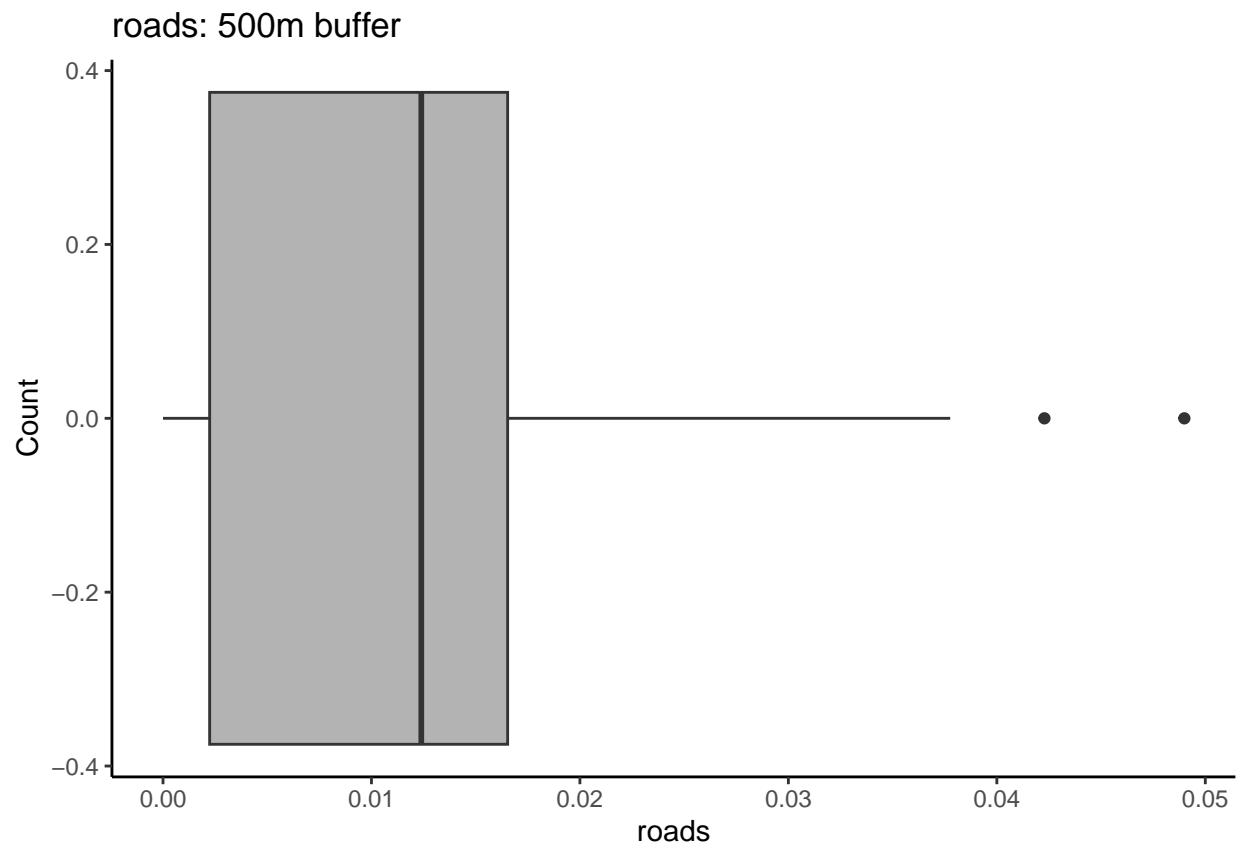




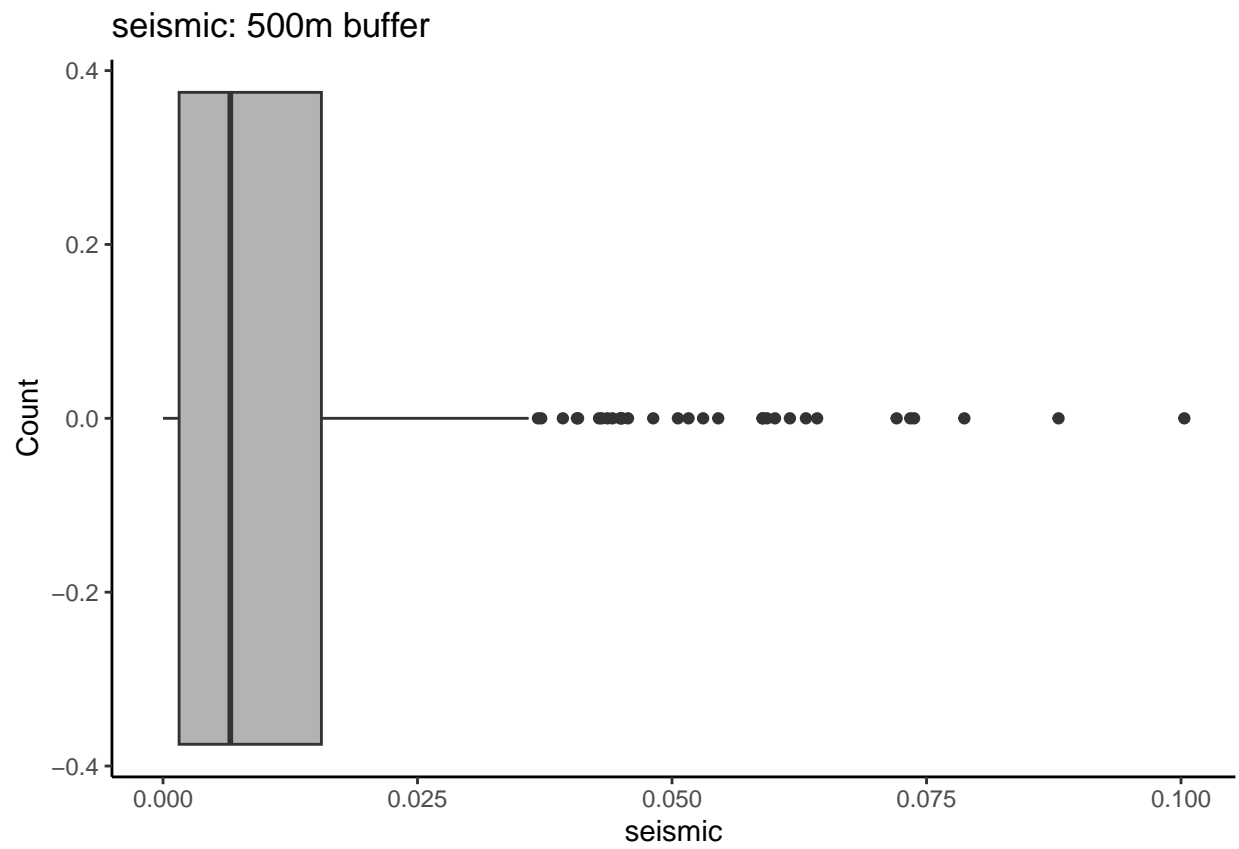
```
##  
## $pipe_trans
```



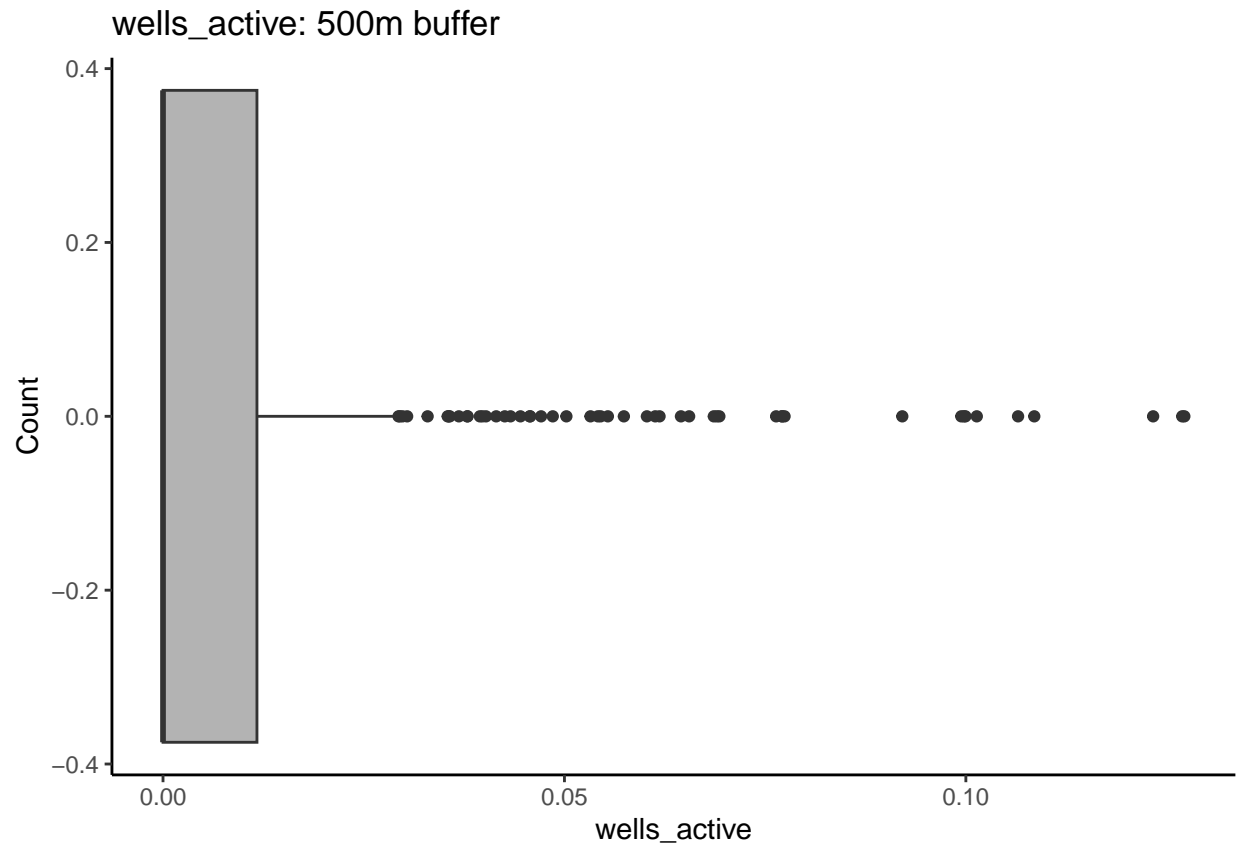
##  
## \$roads



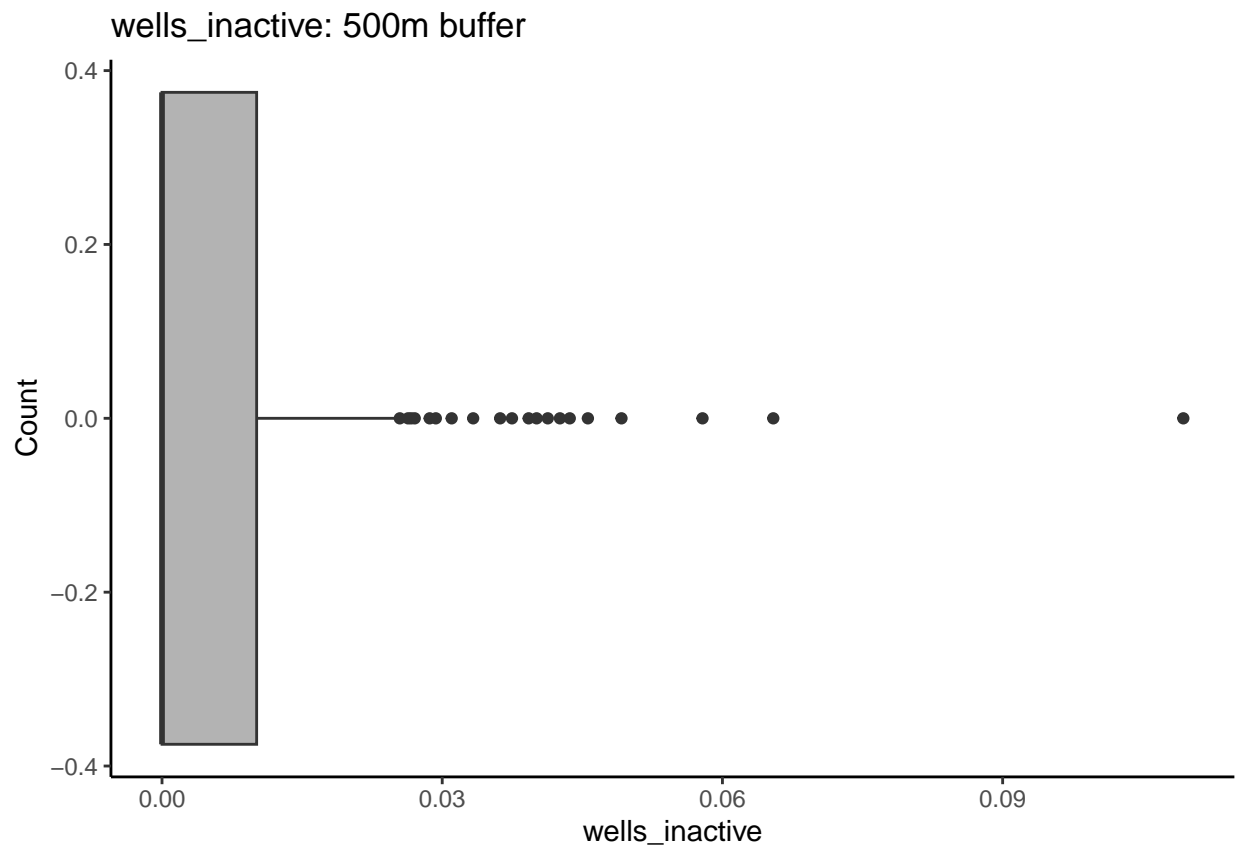
```
##  
## $seismic
```



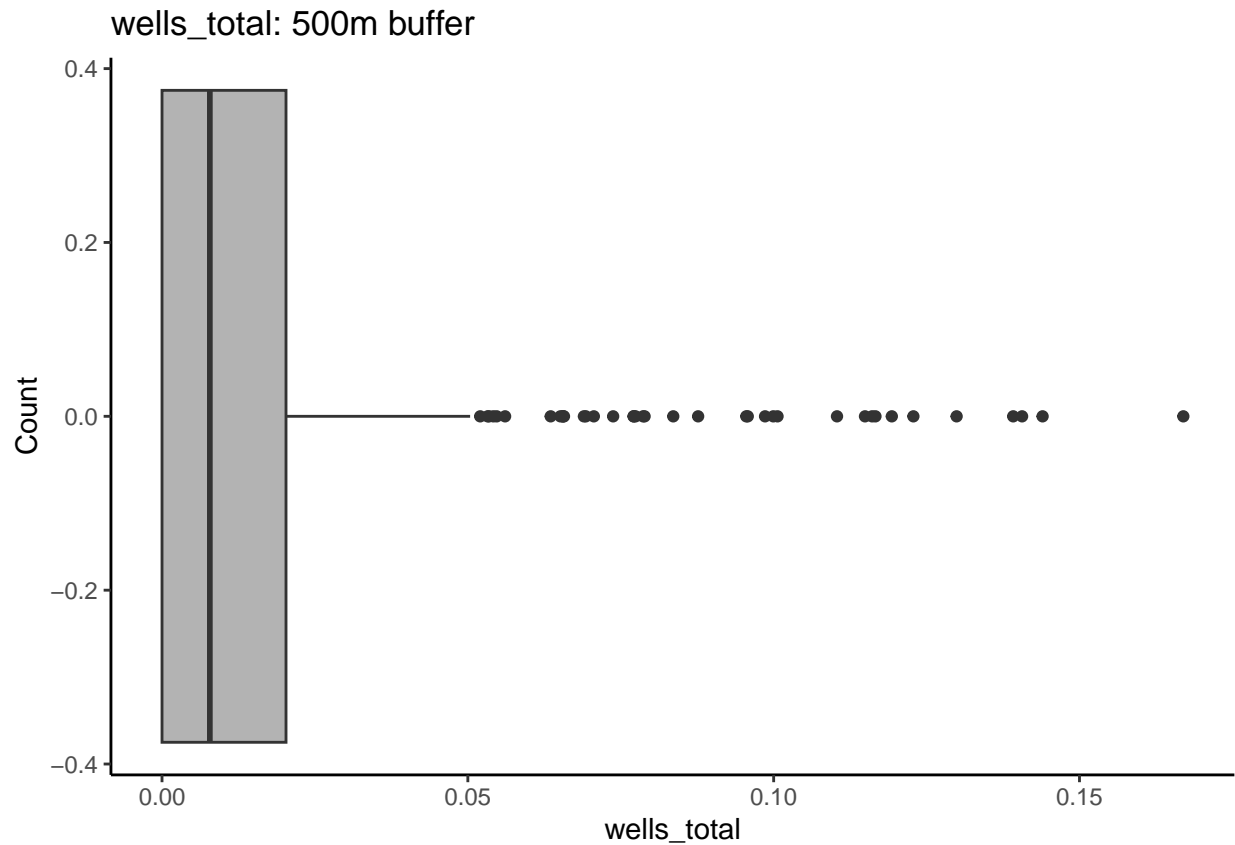
```
##  
## $wells_active
```



```
##  
## $wells_inactive
```



```
##  
## $wells_total
```

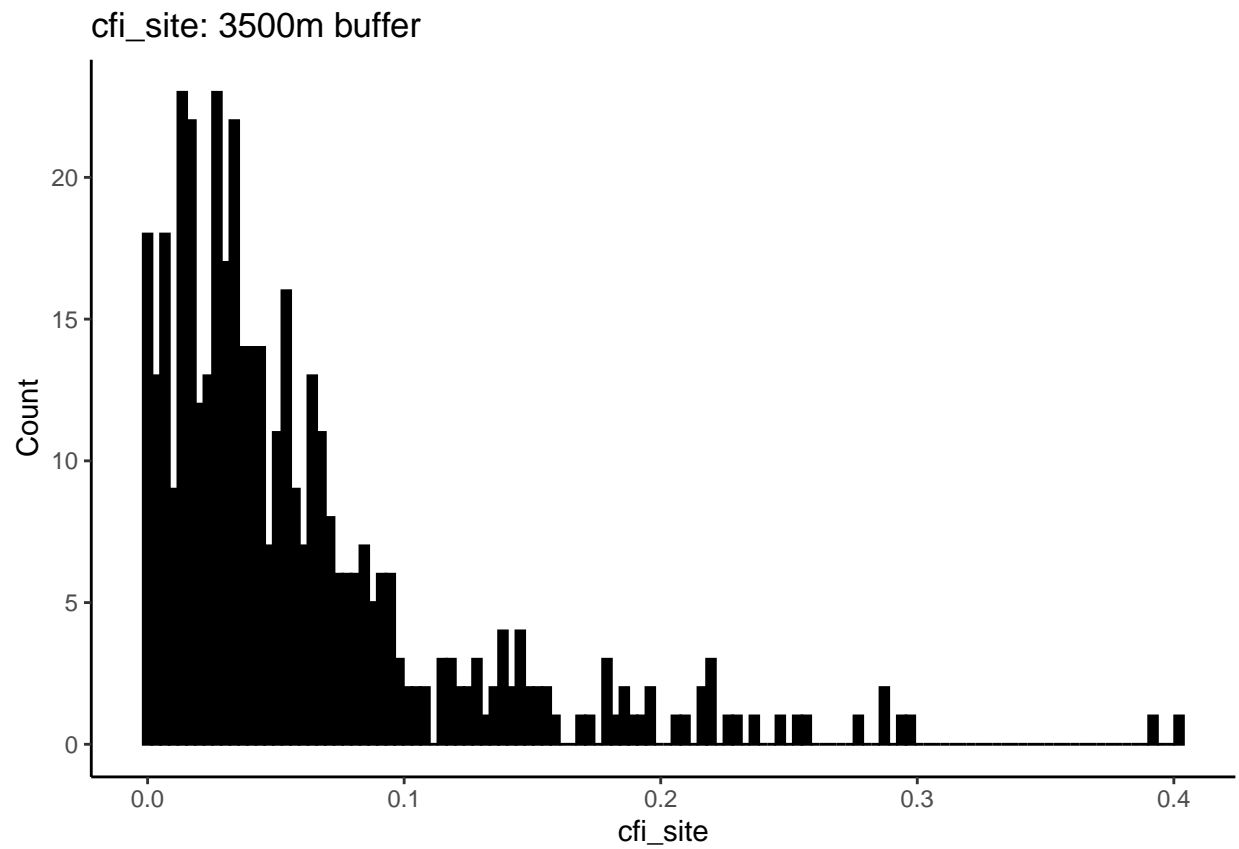


### 2.3. Histograms at 3500 meters:

```
covs_to_explore %>%
  filter(buffer_dist == 3500) %>%
  select(-(1:8)) %>%

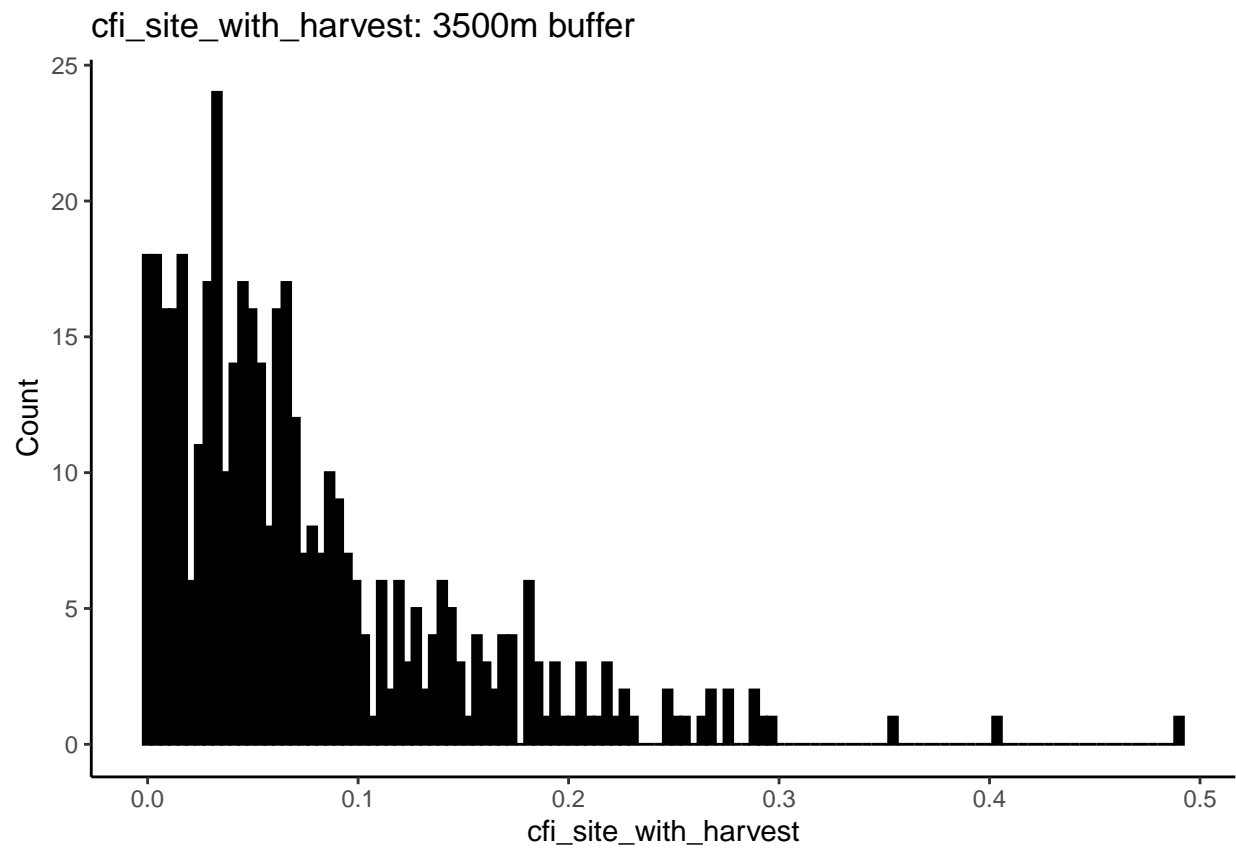
  # use imap which will retain both the data (x) and the variable names (y)
  imap(~ ggplot(mapping = aes(x = .x)) +
    geom_histogram(bins = 120, fill = "black", color = "black") +
    labs(title = paste0(.y, ": 3500m buffer"), x = .y, y = "Count") +
    theme_classic())

## $cfi_site
```

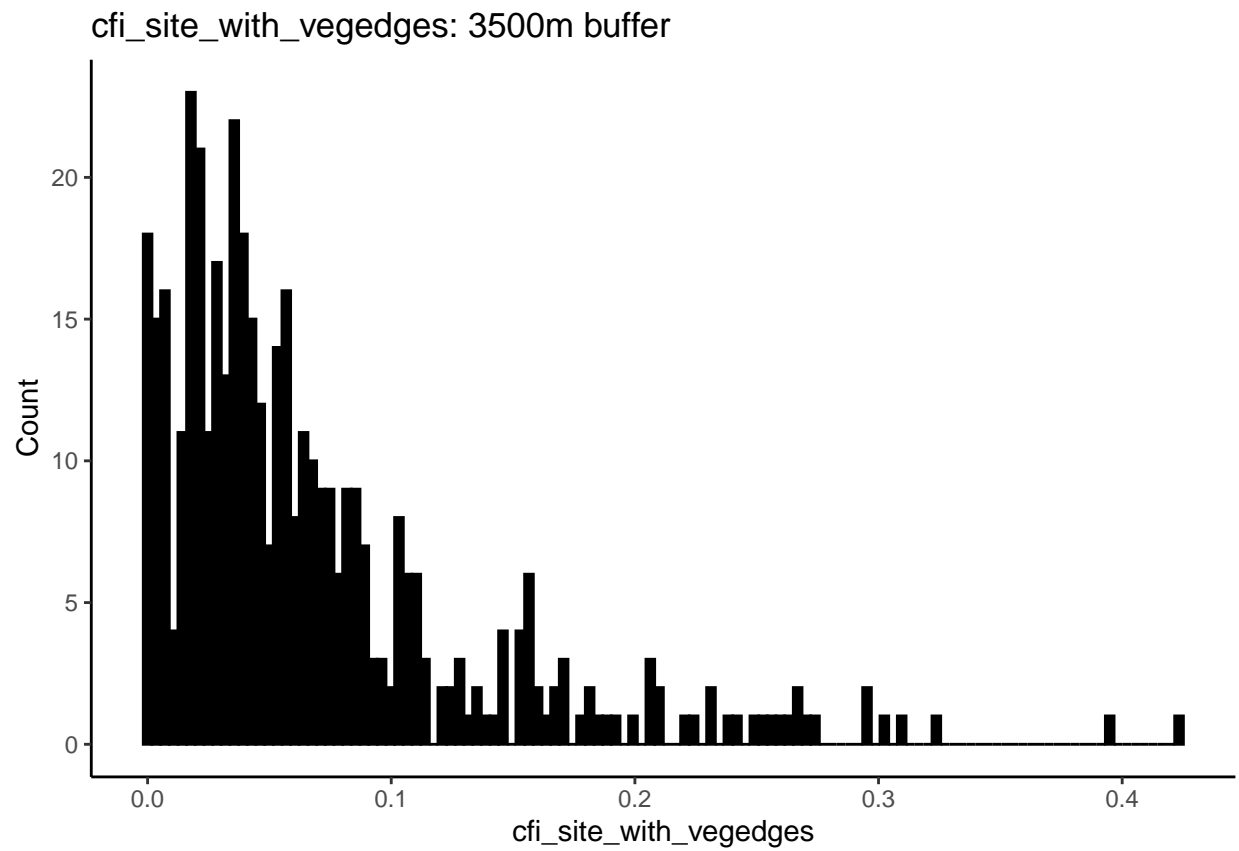


```
##  
## $cfi_site_with_harvest
```

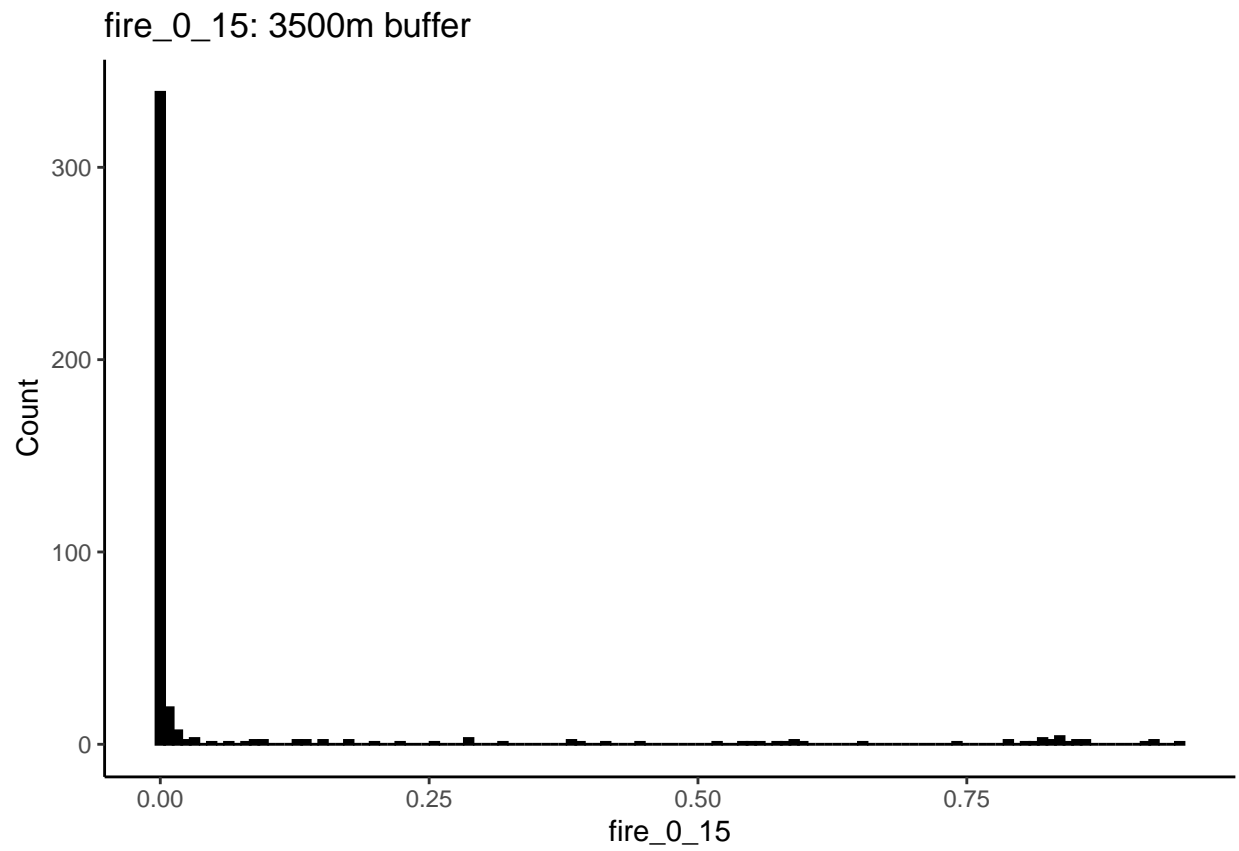




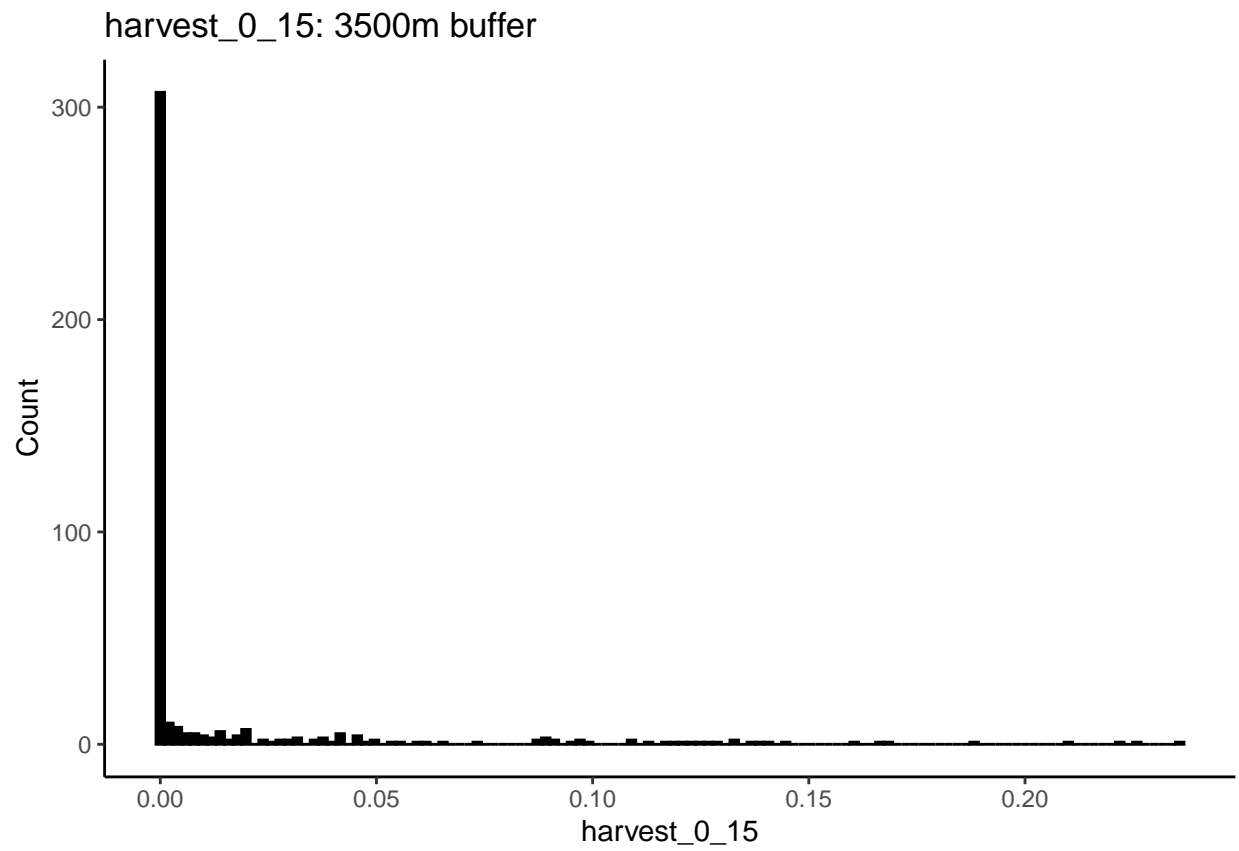
```
##  
## $cfi_site_with_vegedges
```



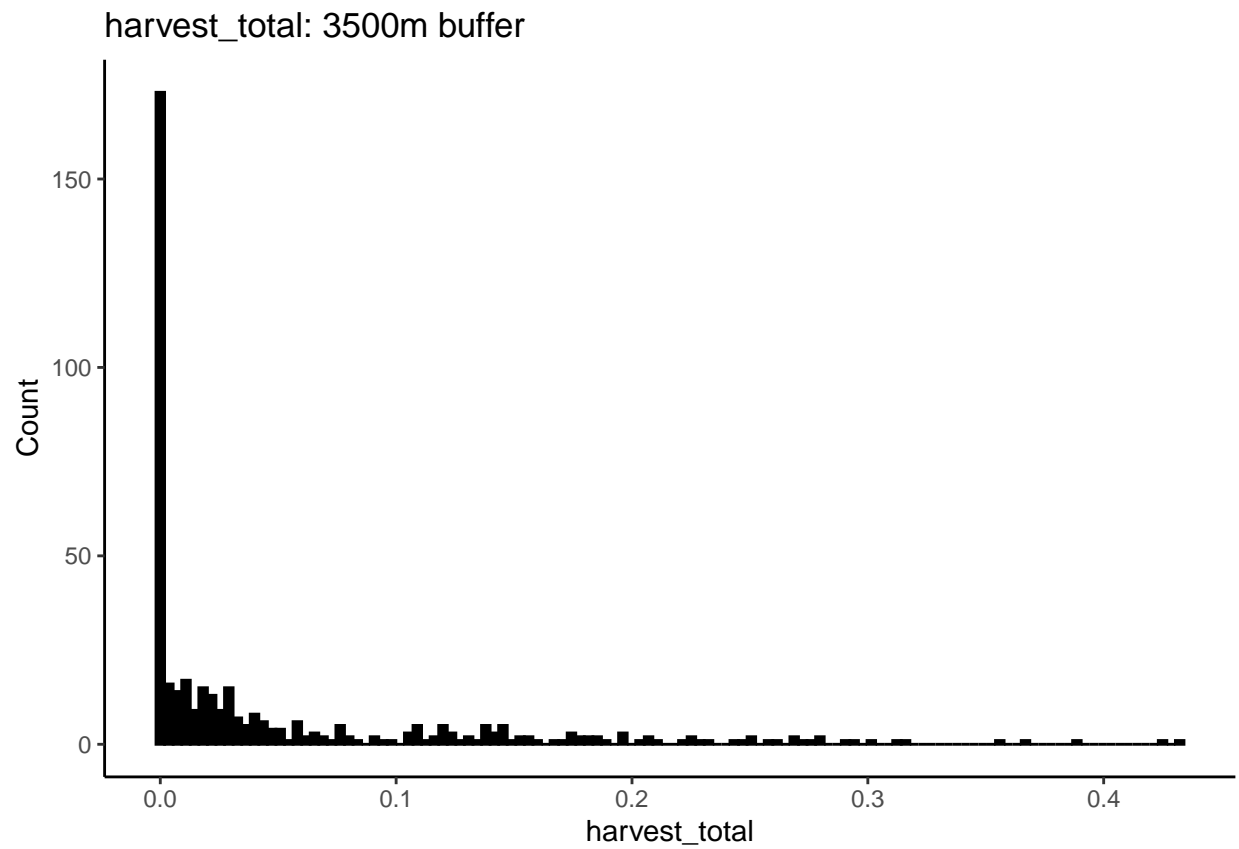
```
##  
## $fire_0_15
```



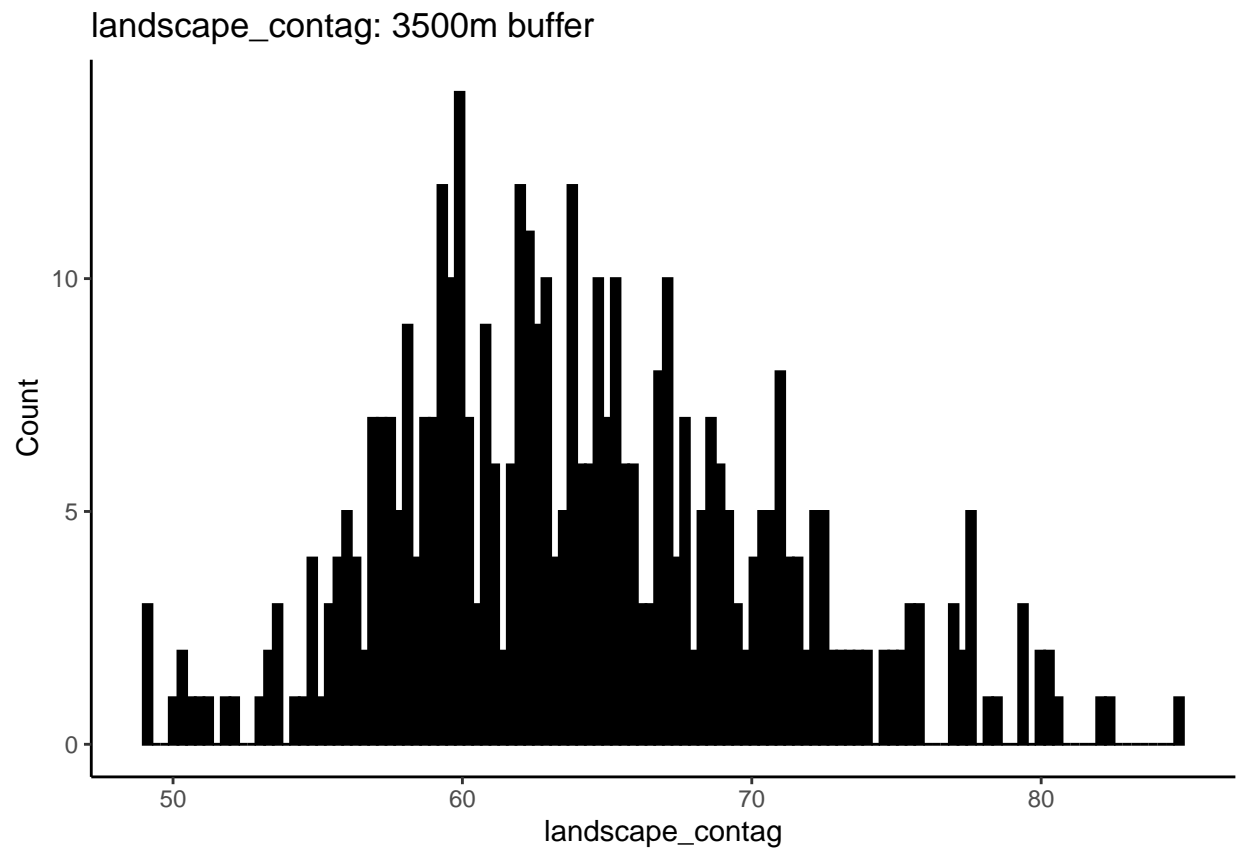
```
##  
## $harvest_0_15
```



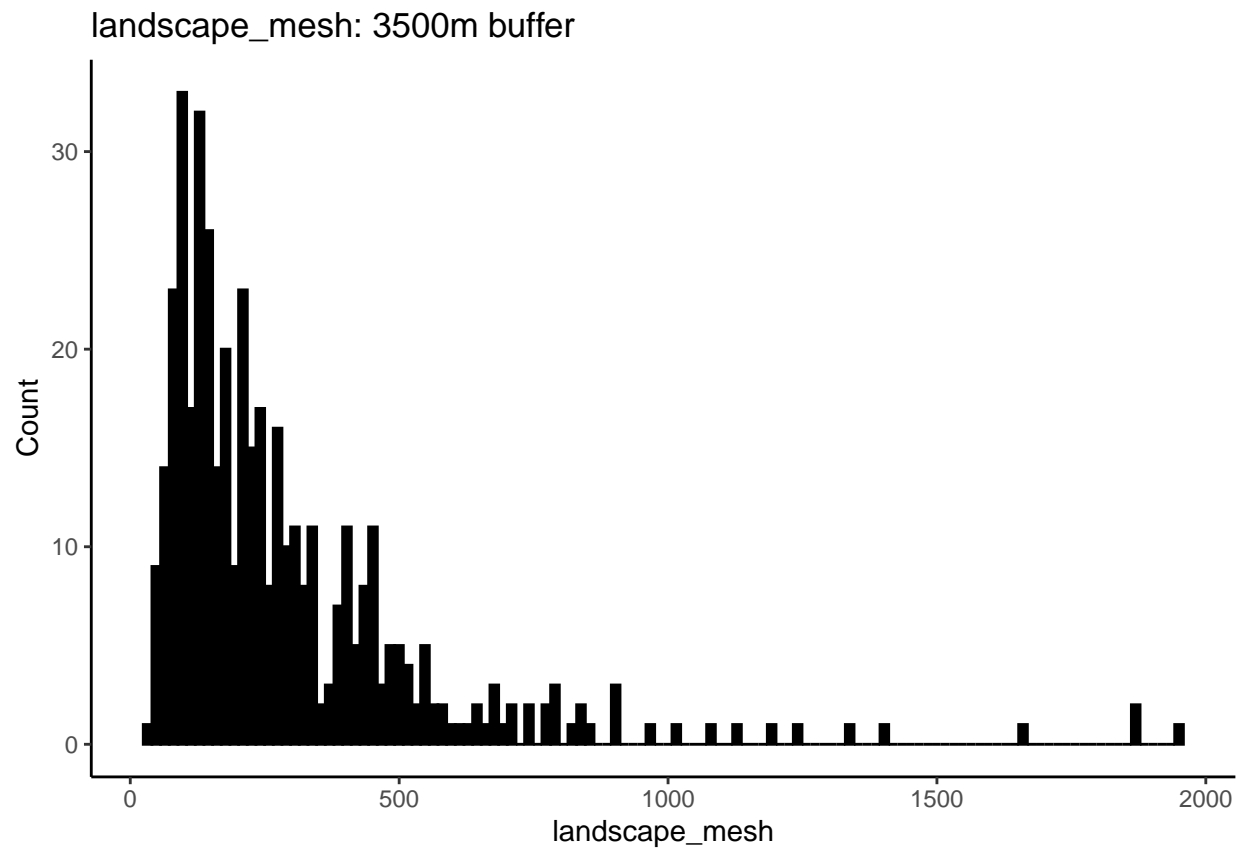
```
##  
## $harvest_total
```



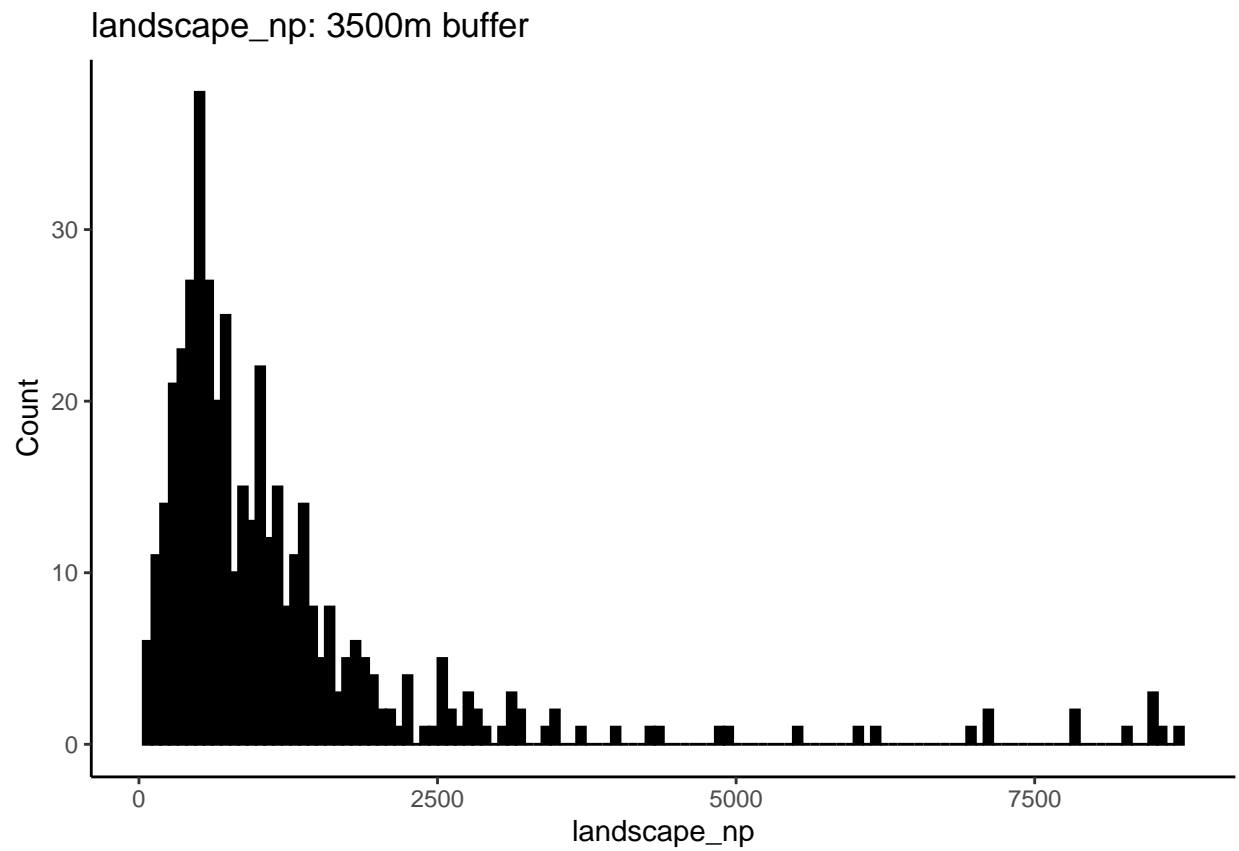
```
##  
## $landscape_contag
```



```
##  
## $landscape_mesh
```

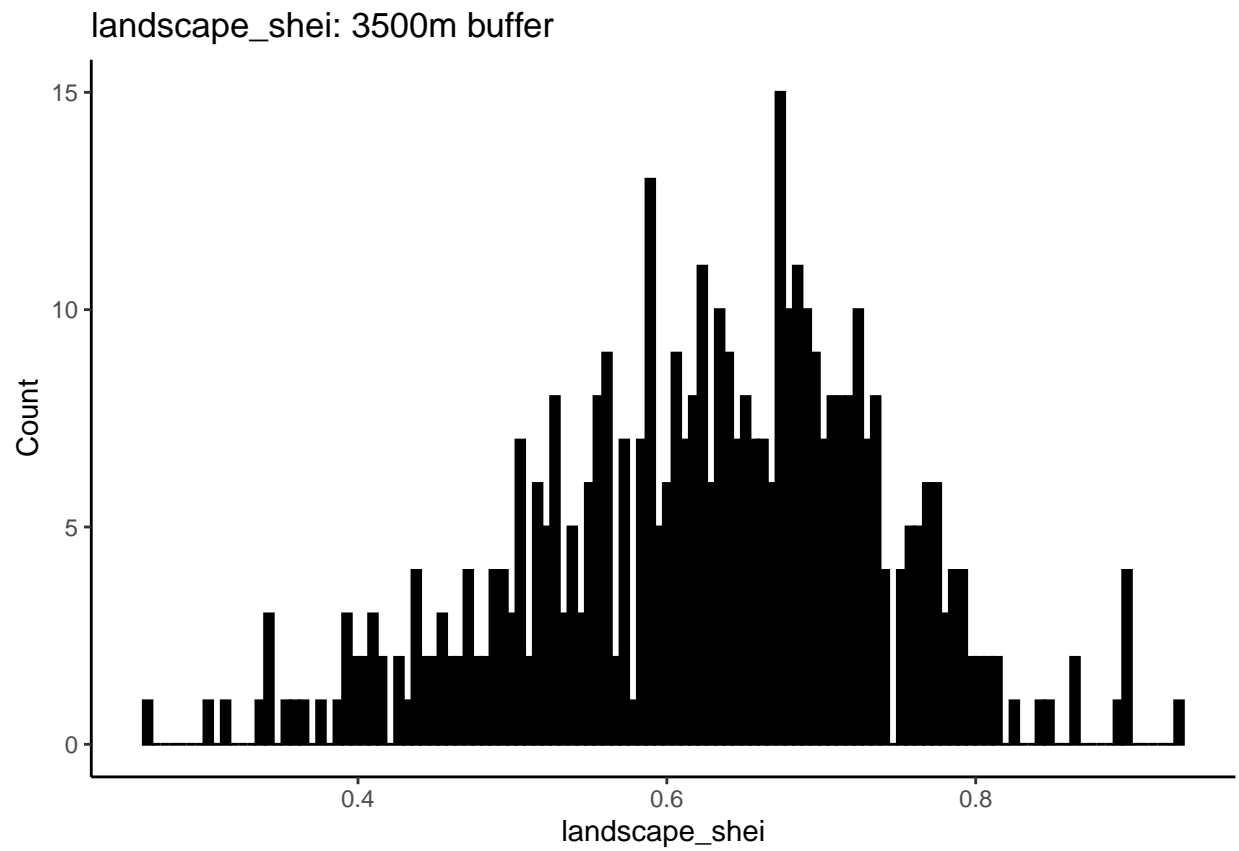


```
##  
## $landscape_np
```

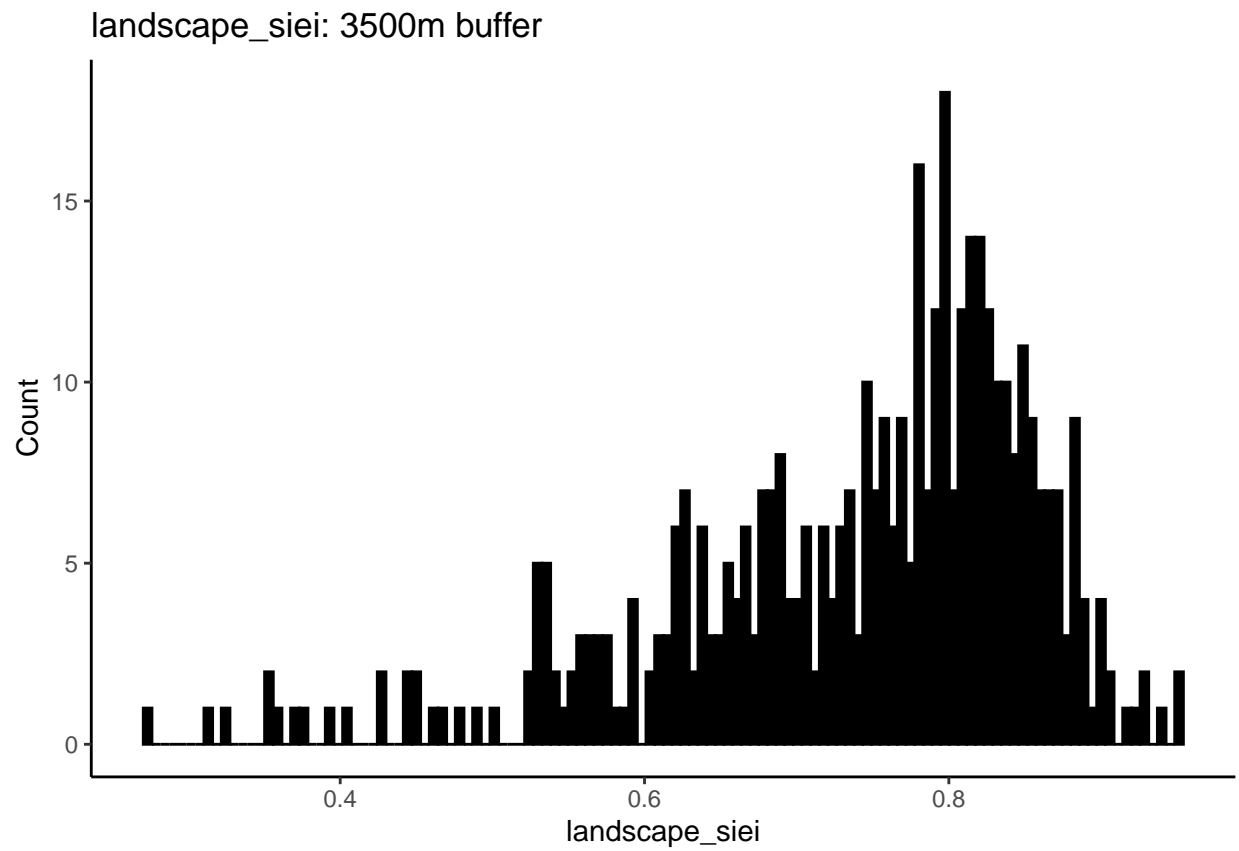


```
##  
## $landscape_shei
```

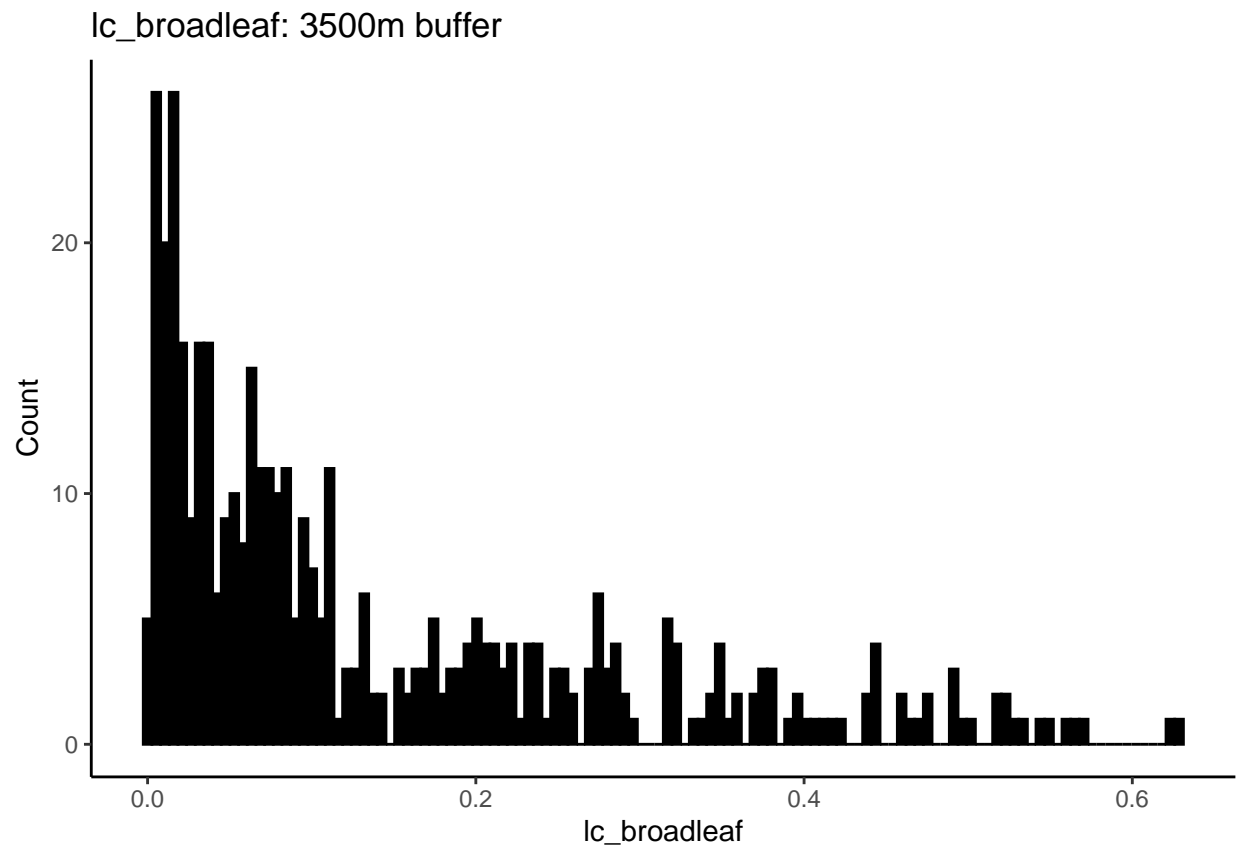




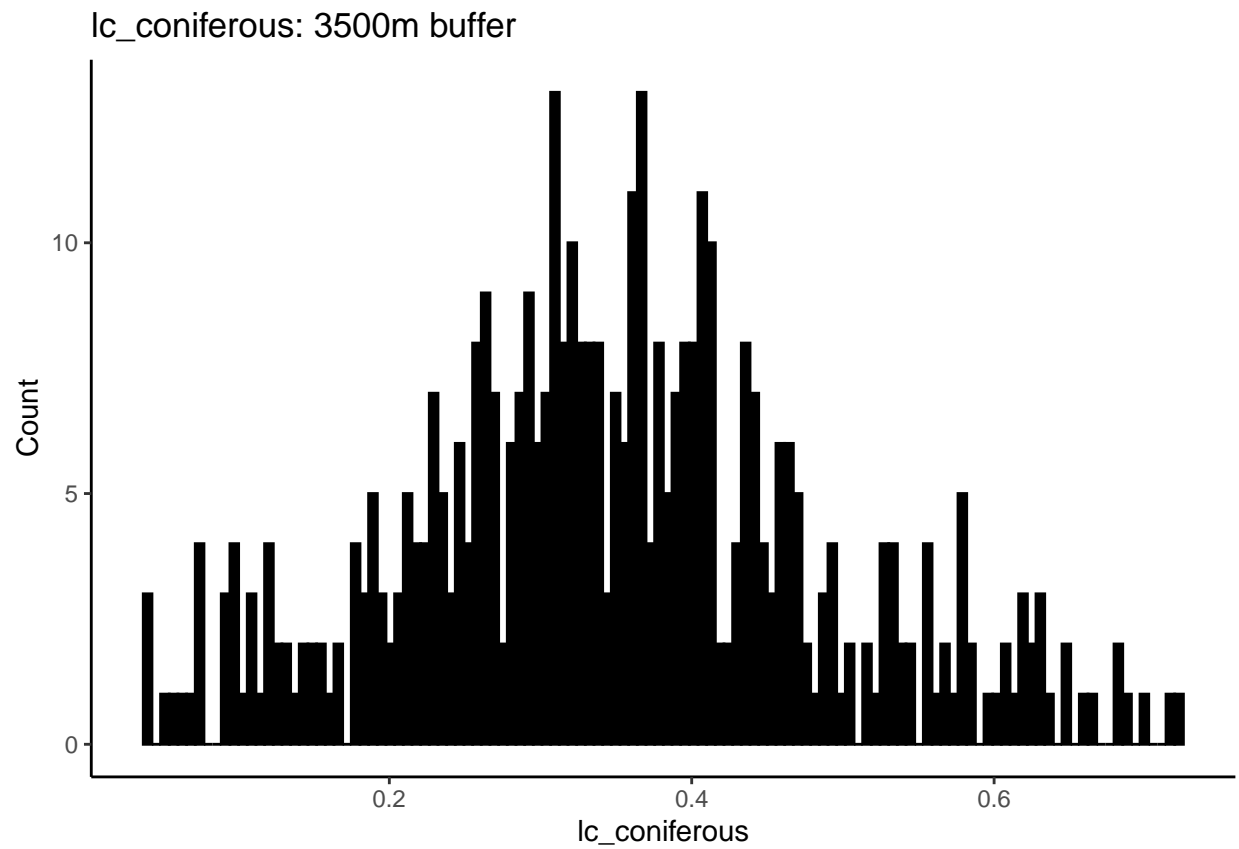
```
##  
## $landscape_siei
```



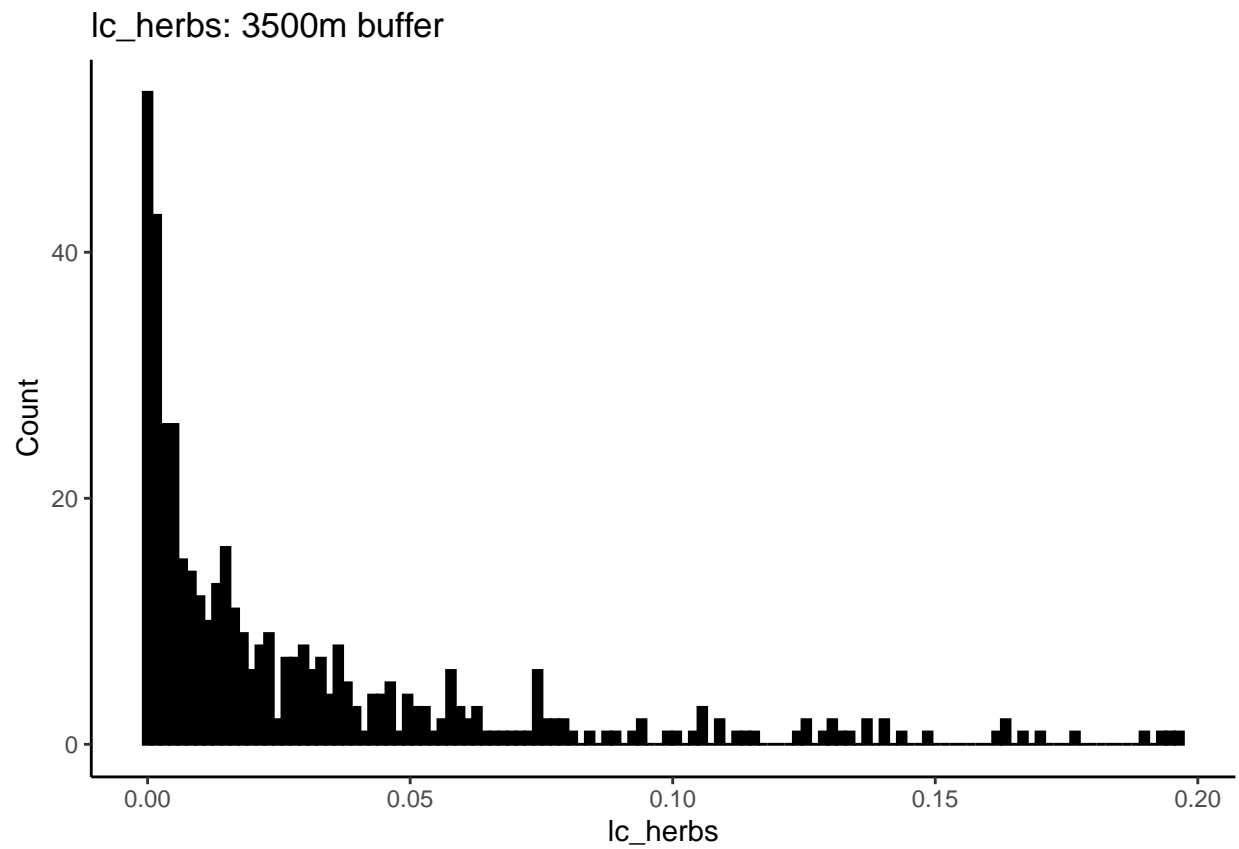
```
##  
## $lc_broadleaf
```



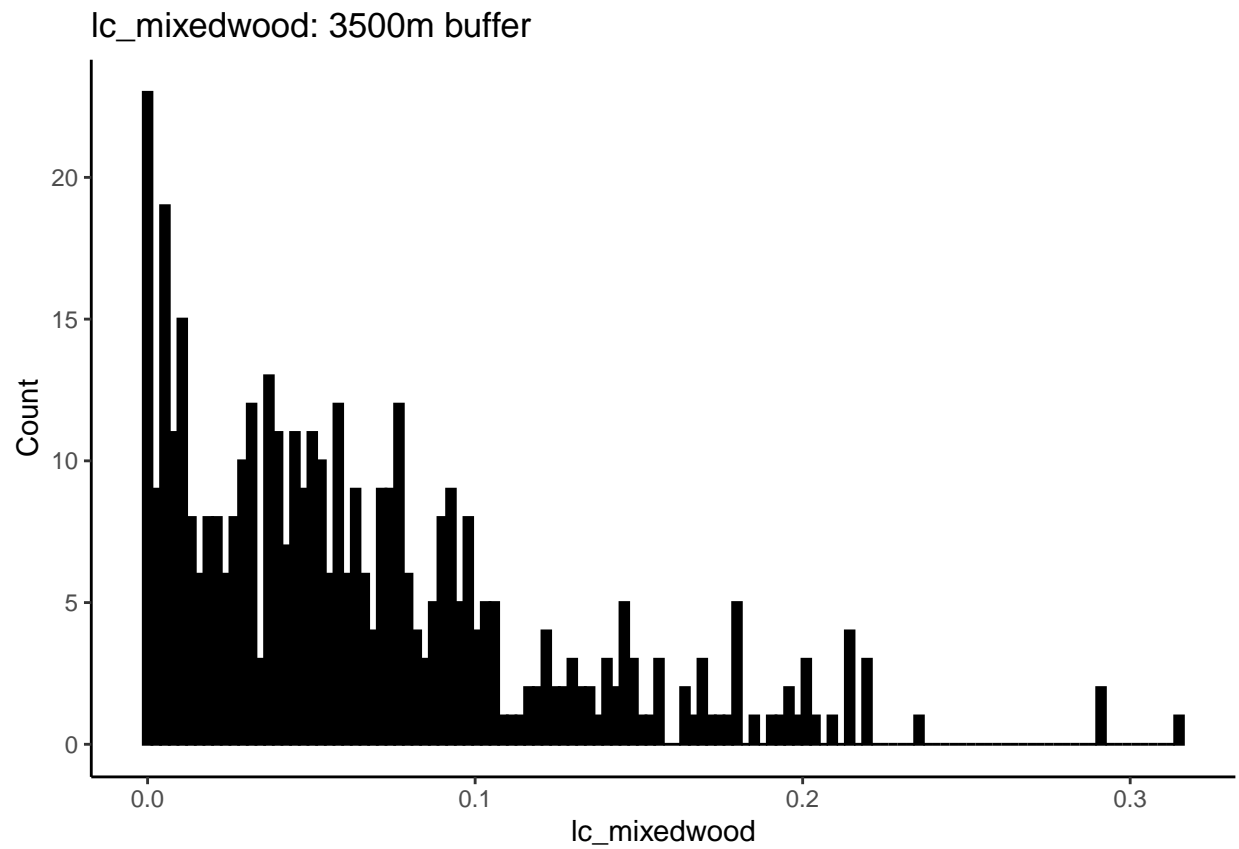
```
##  
## $lc_coniferous
```



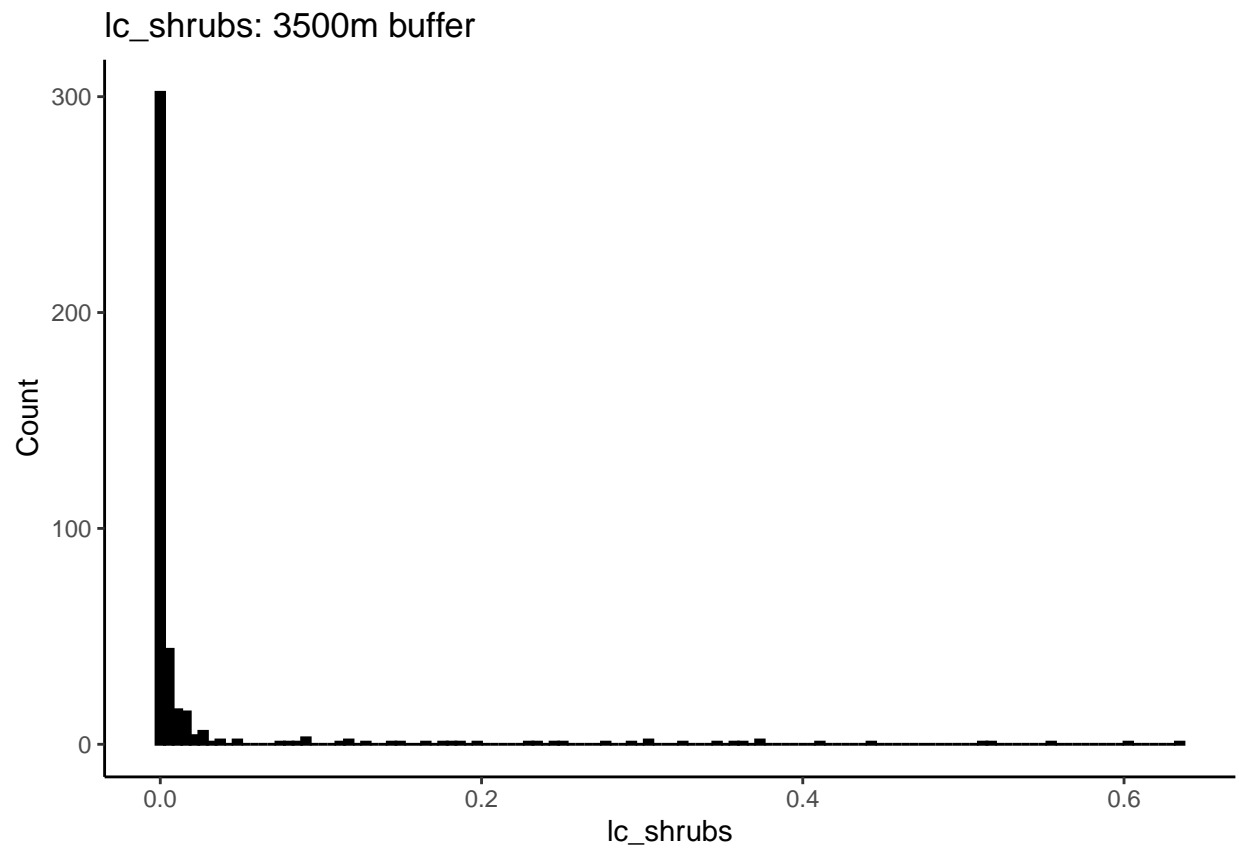
```
##  
## $lc_herbs
```



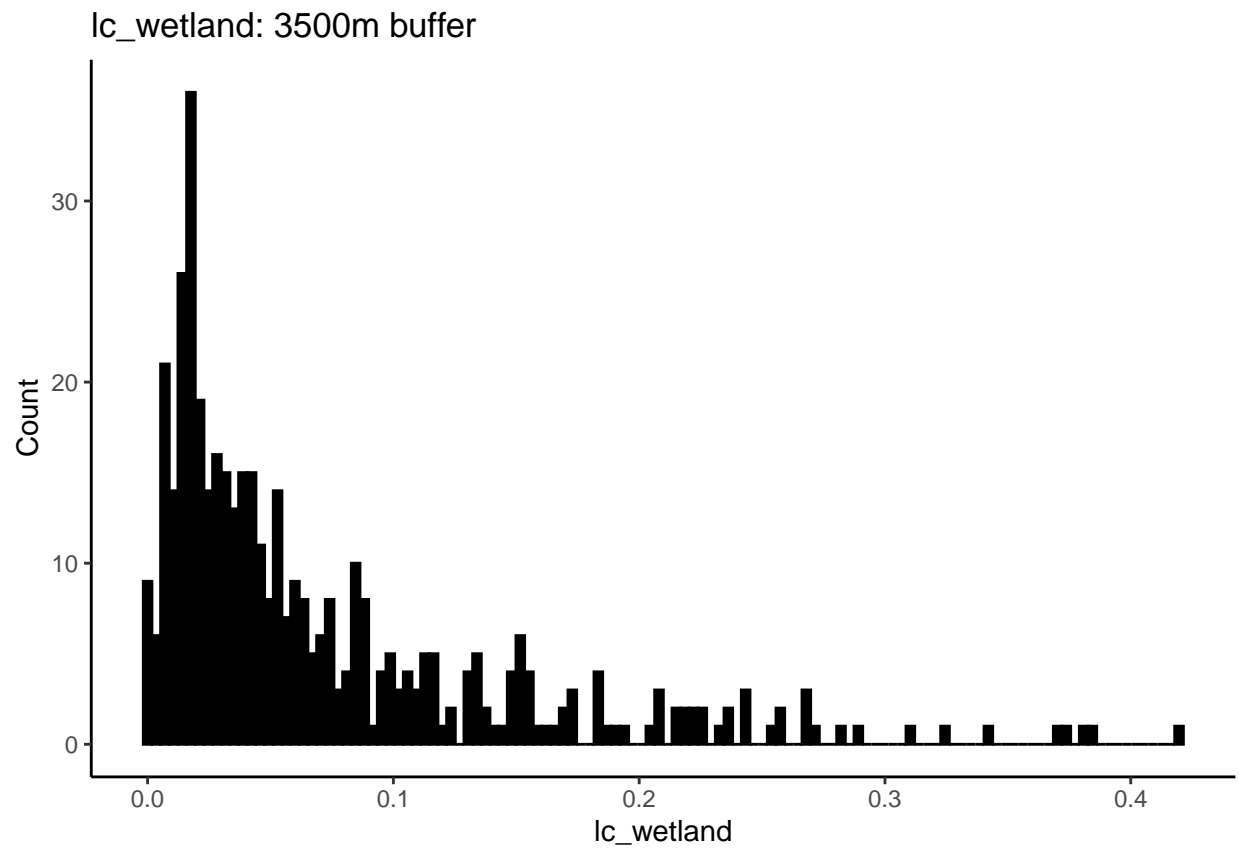
```
##  
## $lc_mixedwood
```



```
##  
## $lc_shrubs
```

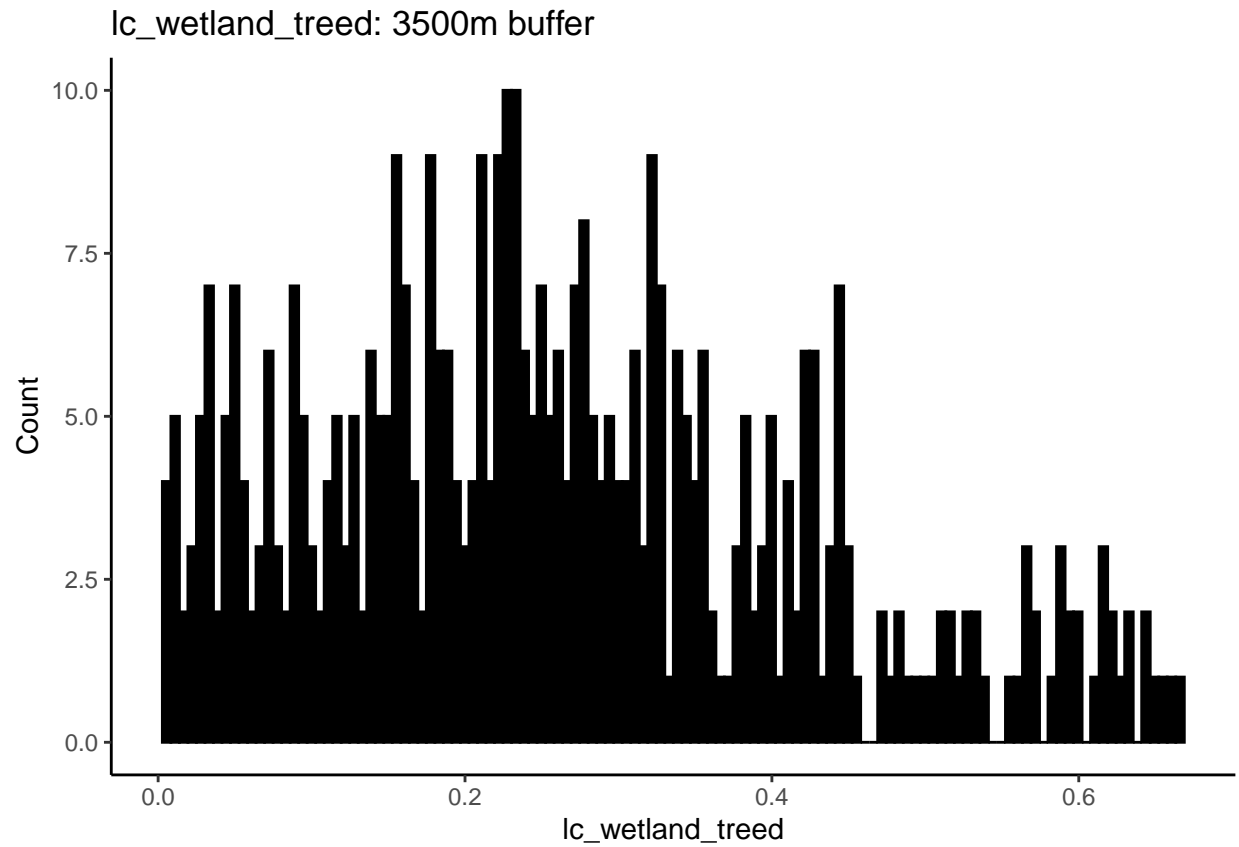


```
##  
## $lc_wetland
```

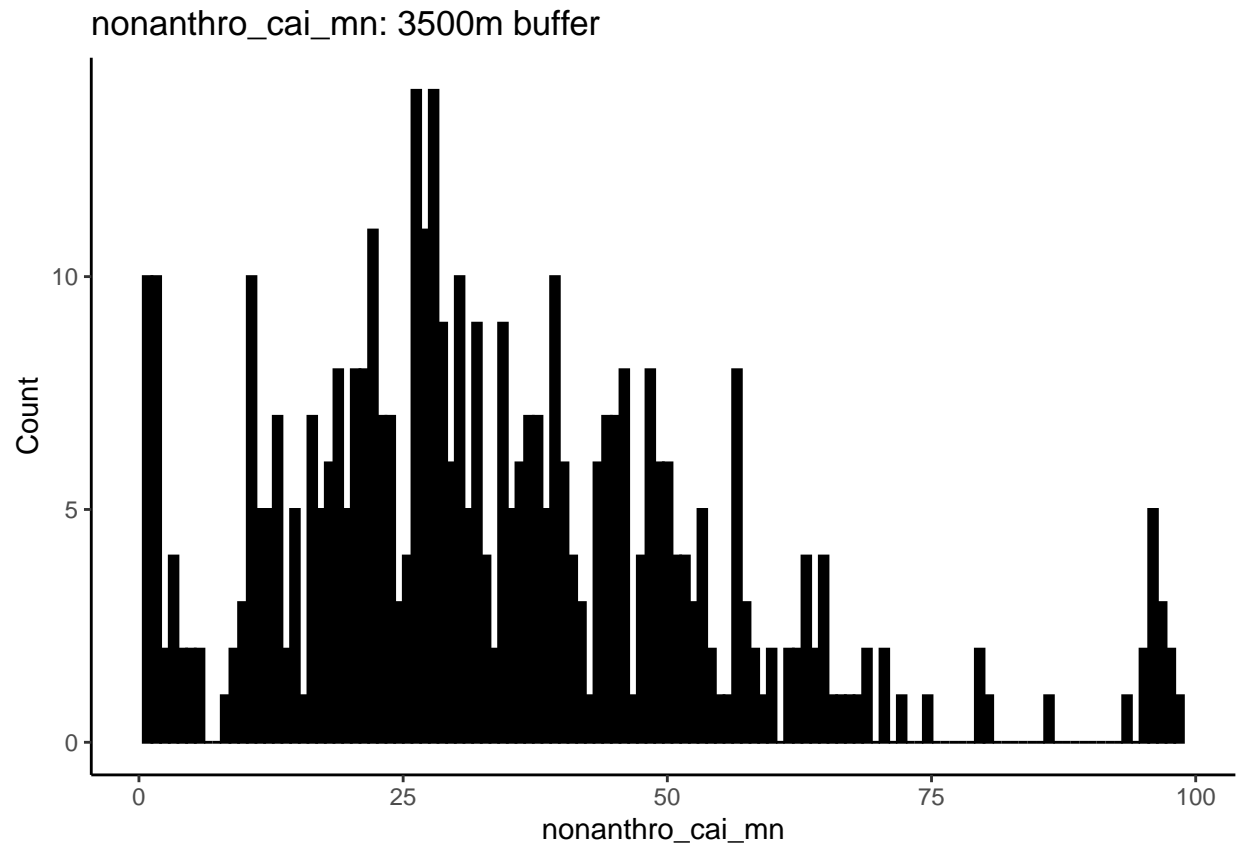


```
##  
## $lc_wetland_treed
```

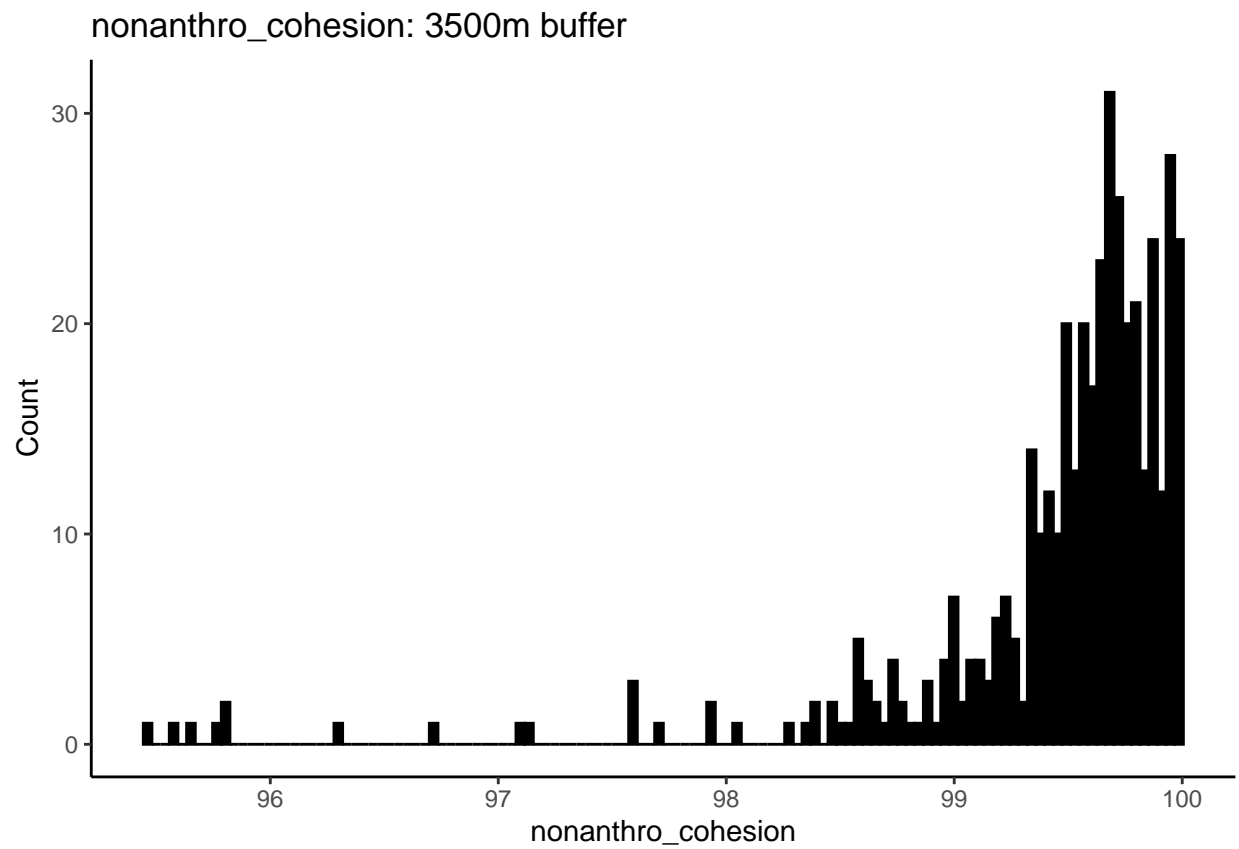




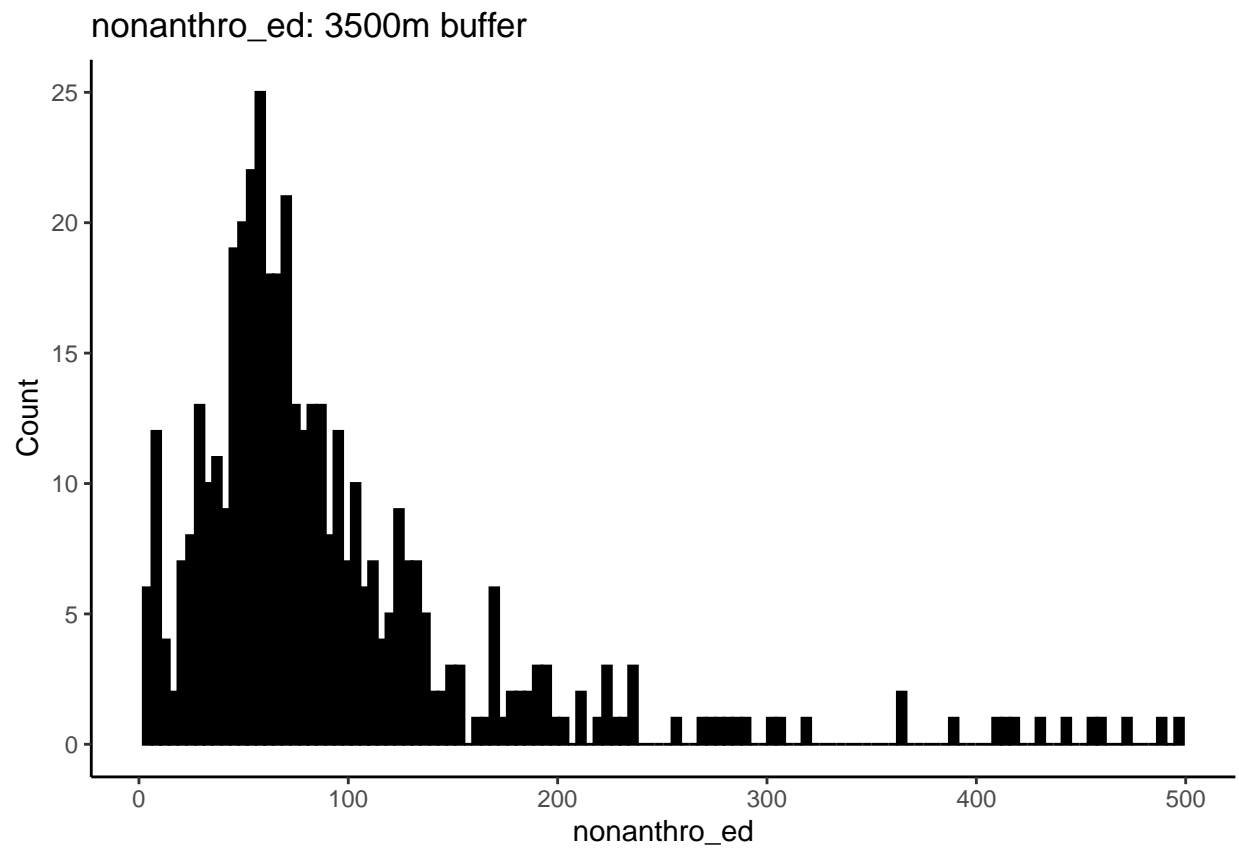
```
##  
## $nonanthro_cai_mn
```



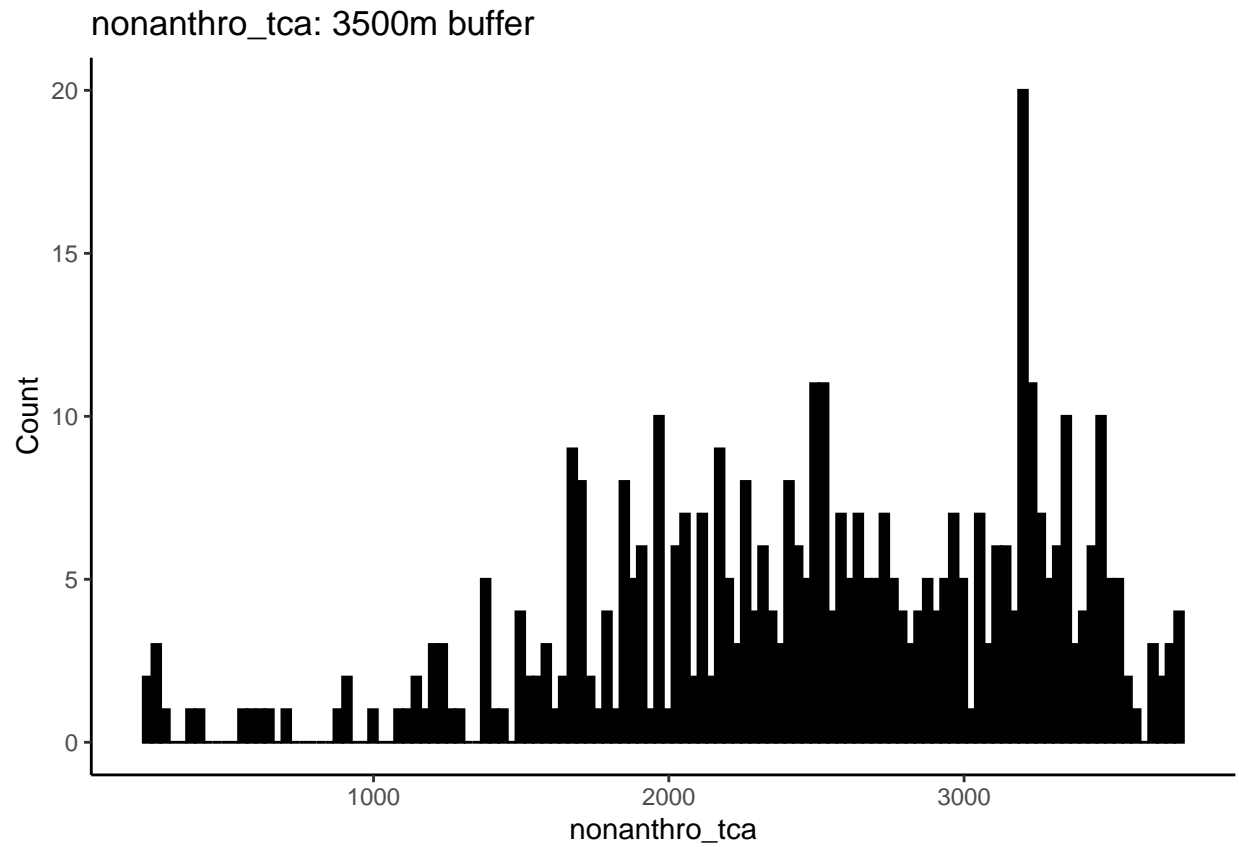
```
##  
## $nonanthro_cohesion
```



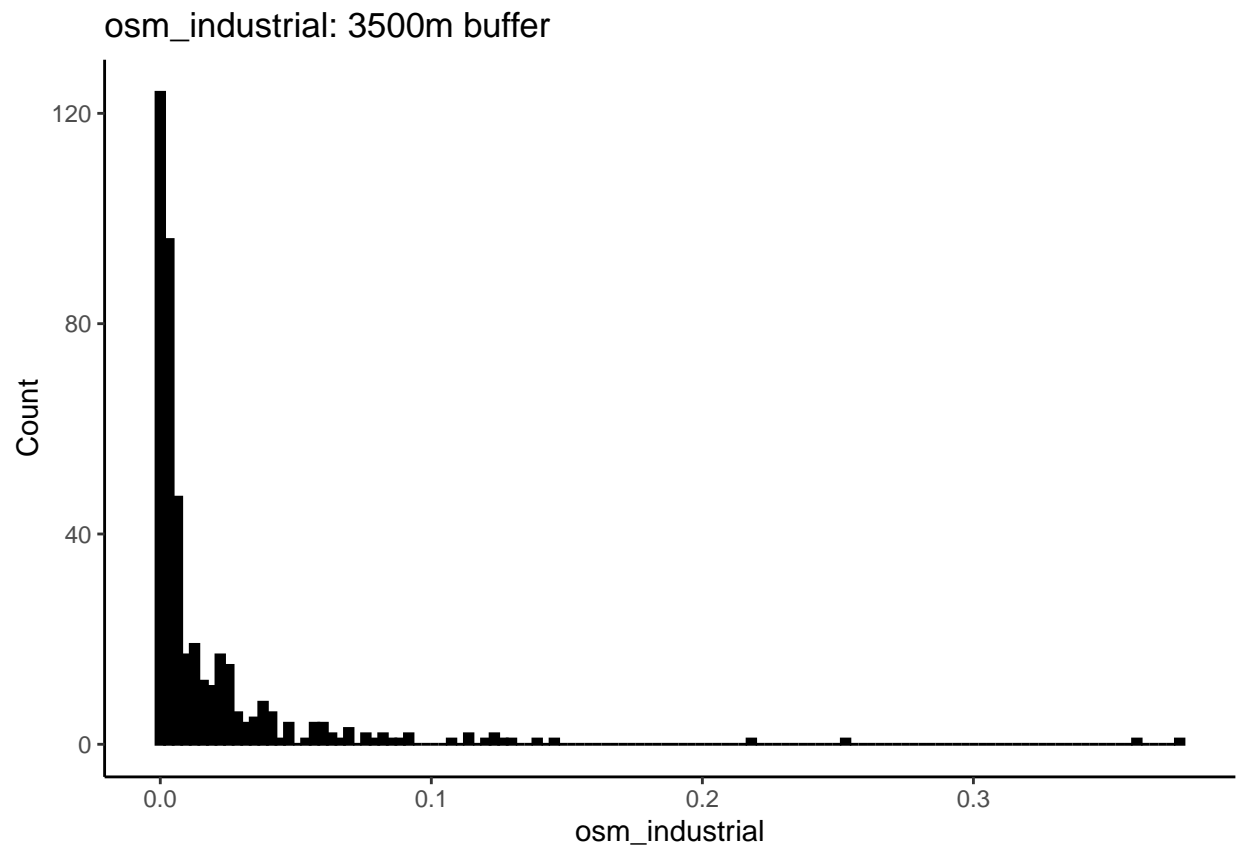
```
##  
## $nonanthro_ed
```



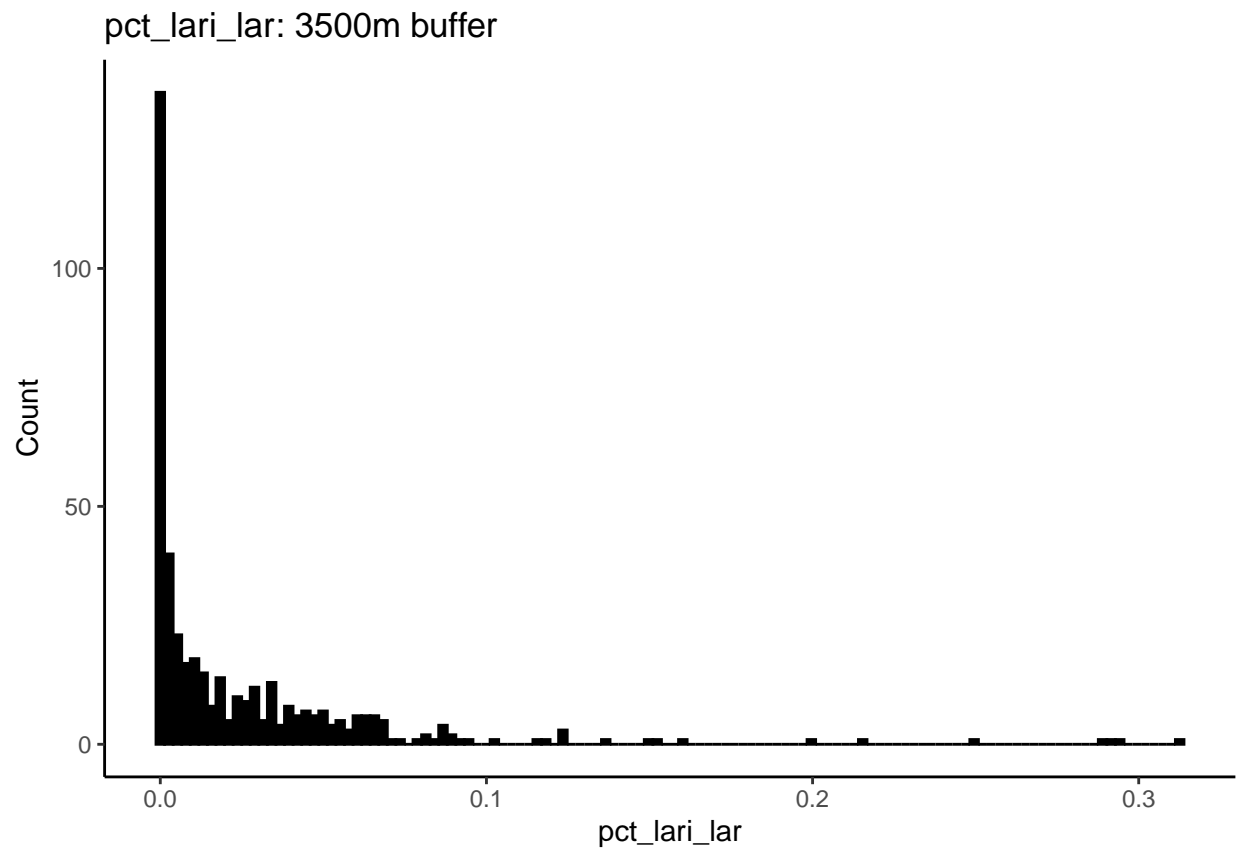
```
##  
## $nonanthro_tca
```



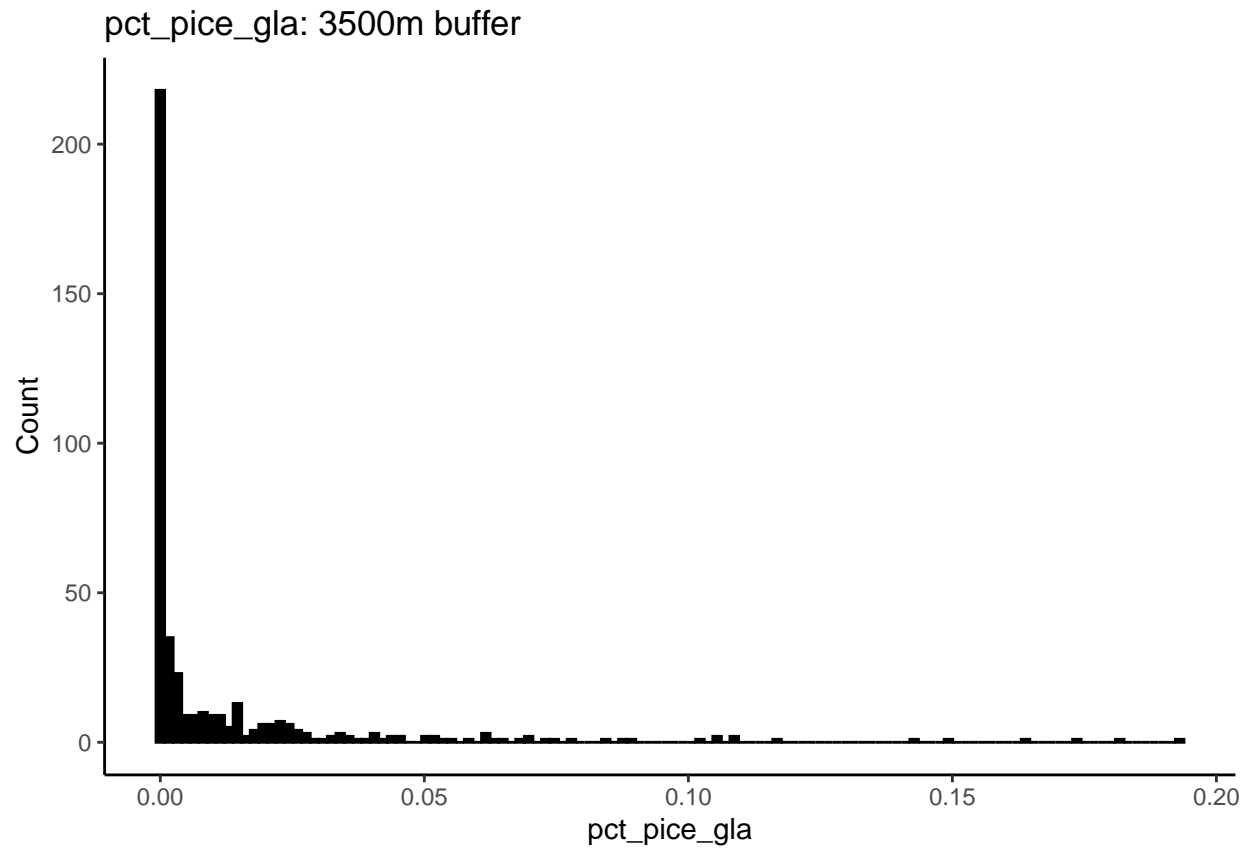
```
##  
## $osm_industrial
```



```
##  
## $pct_lari_lar
```

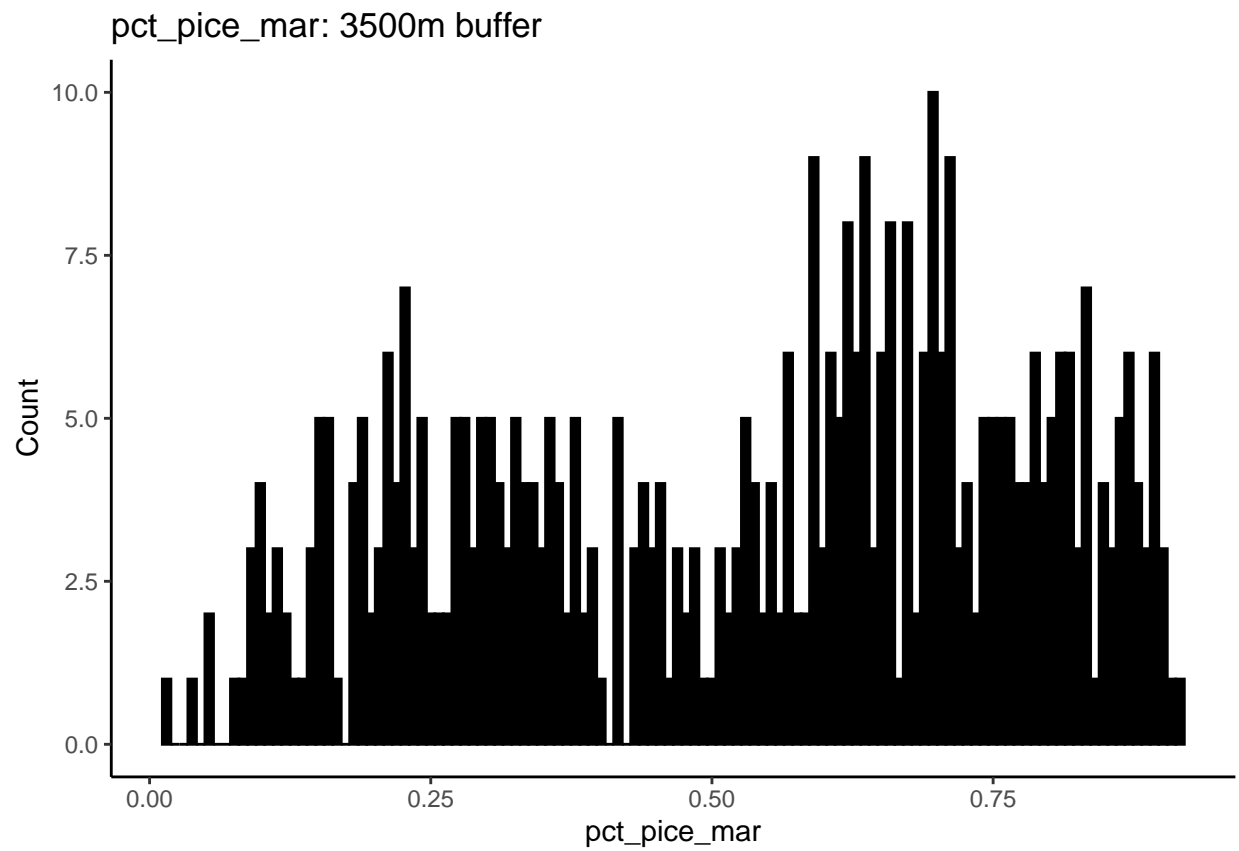


```
##  
## $pct_pice_gla
```

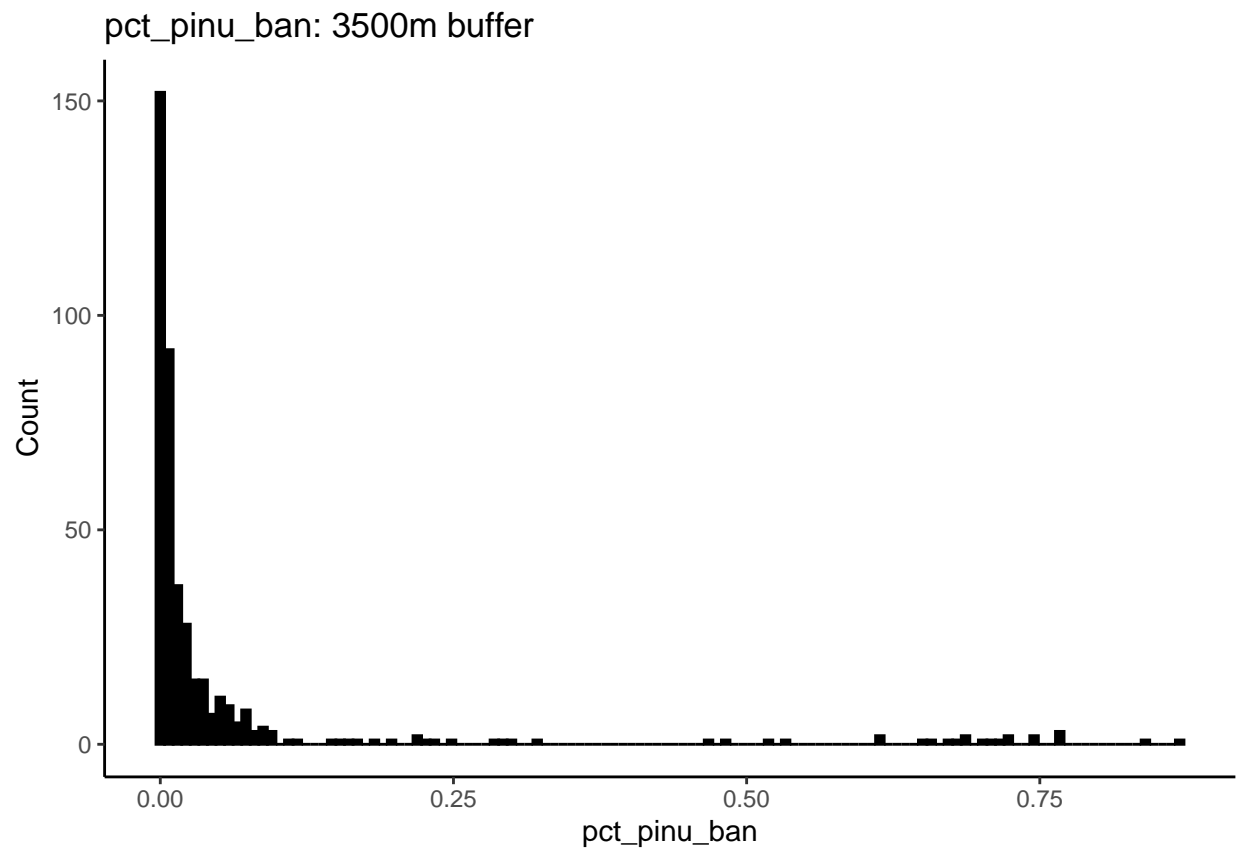


```
##  
## $pct_pice_mar
```

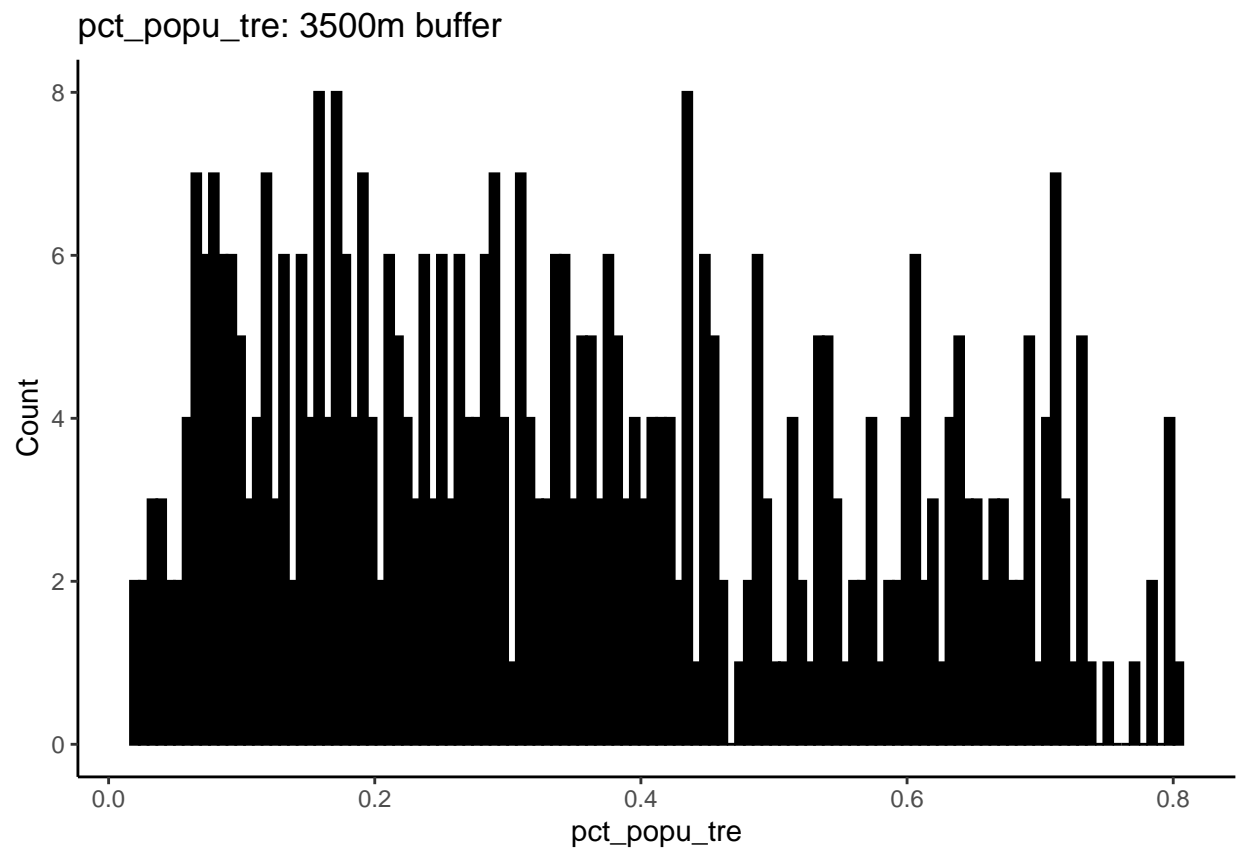




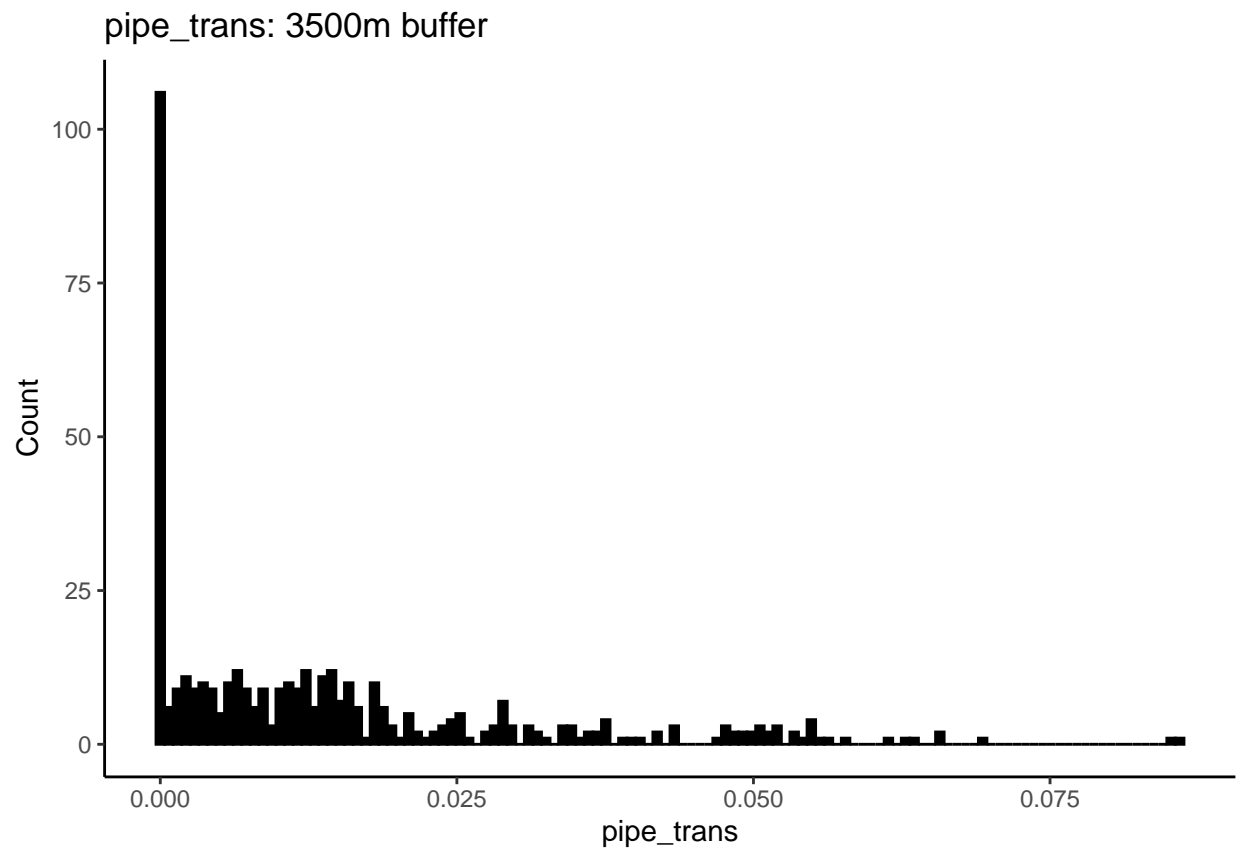
```
##  
## $pct_pinu_ban
```



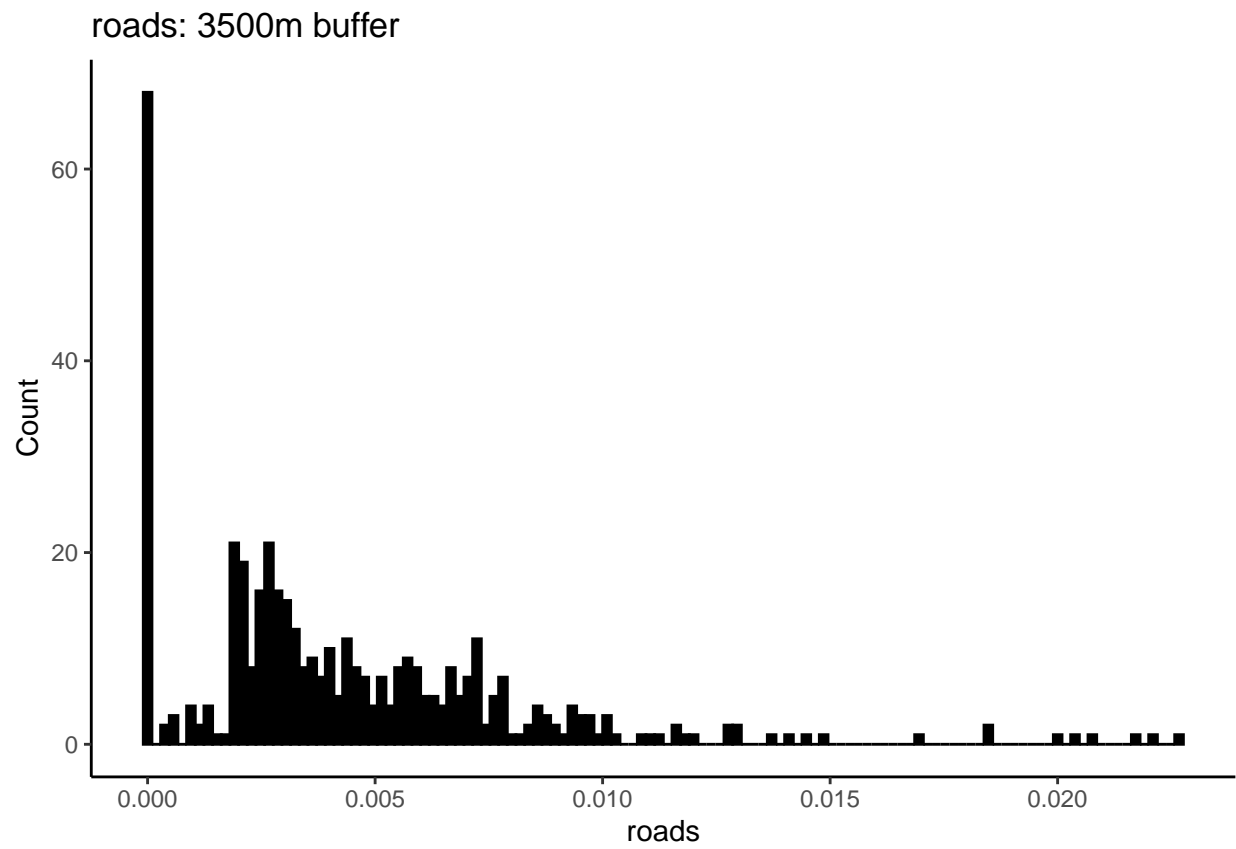
```
##  
## $pct_popu_tre
```



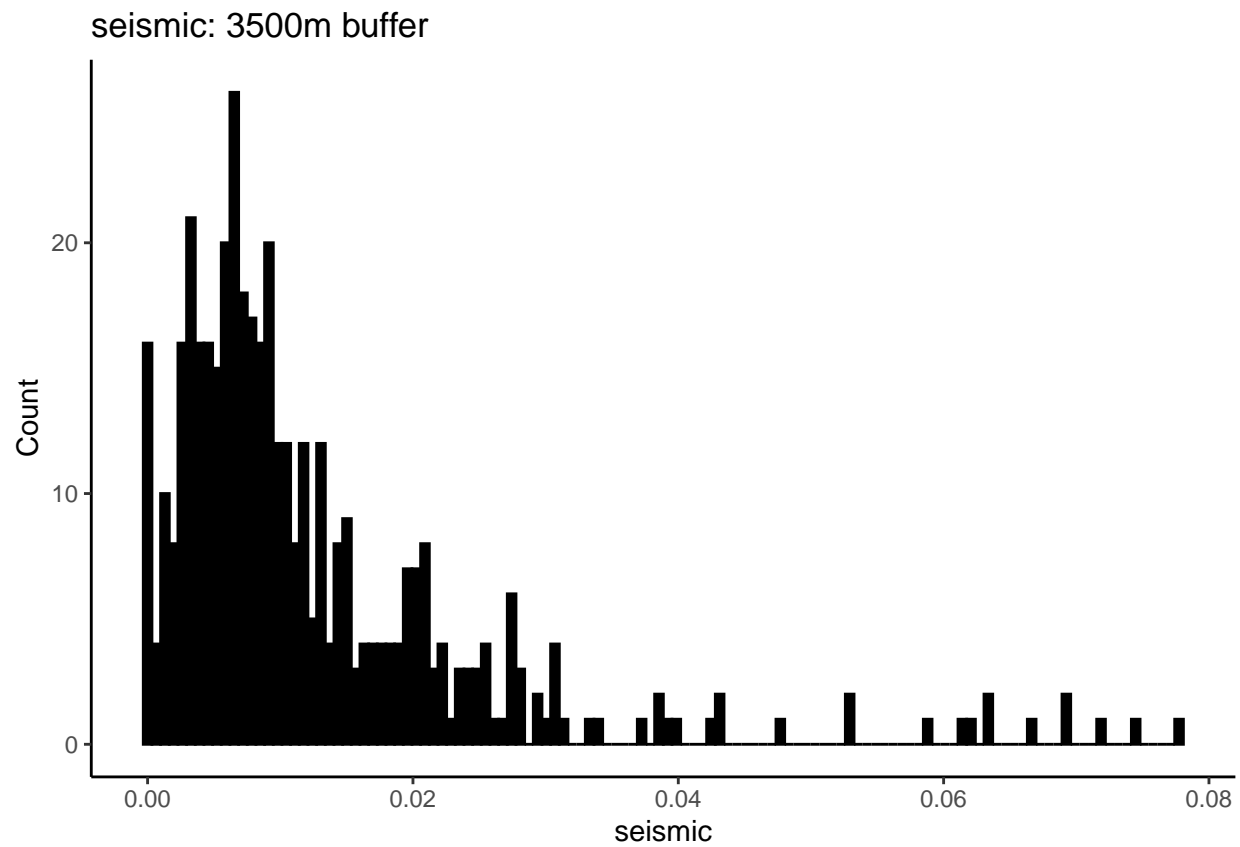
```
##  
## $pipe_trans
```



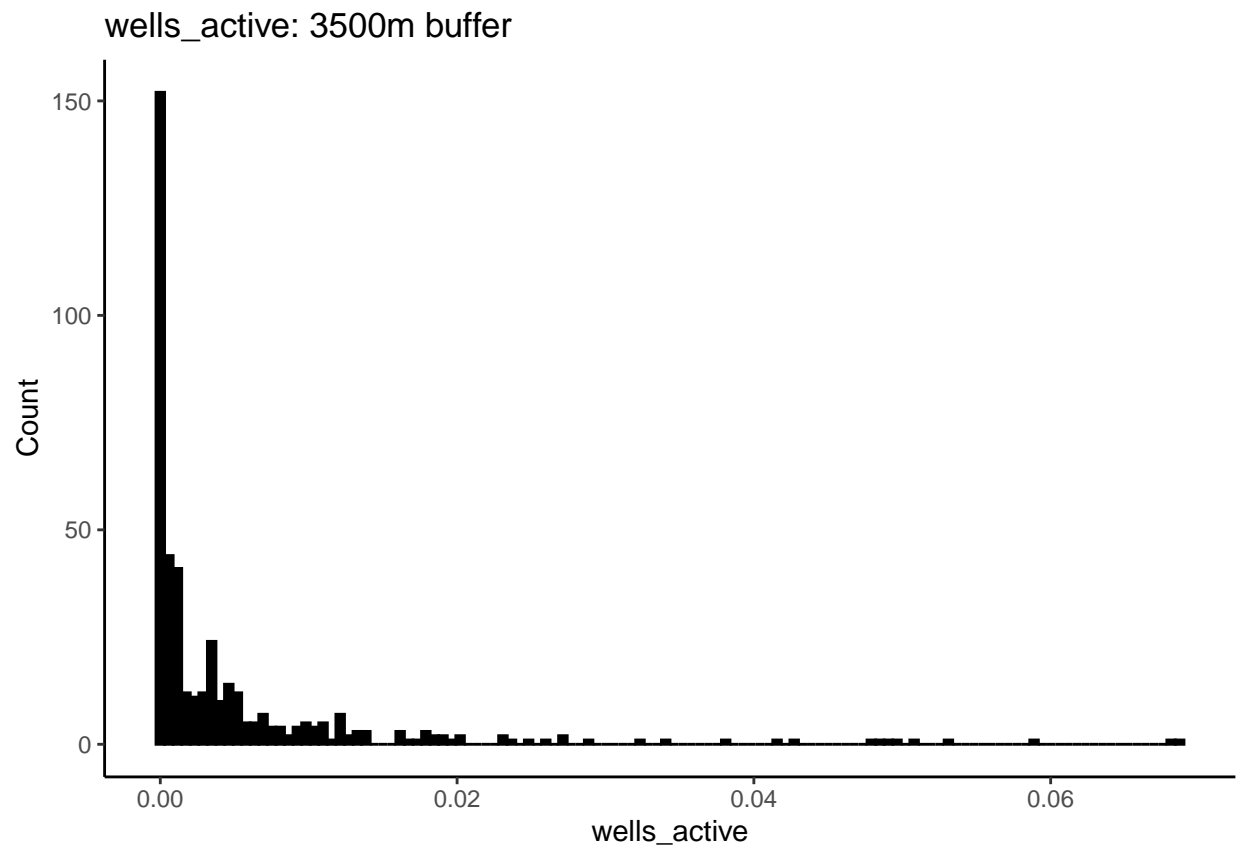
```
##  
## $roads
```



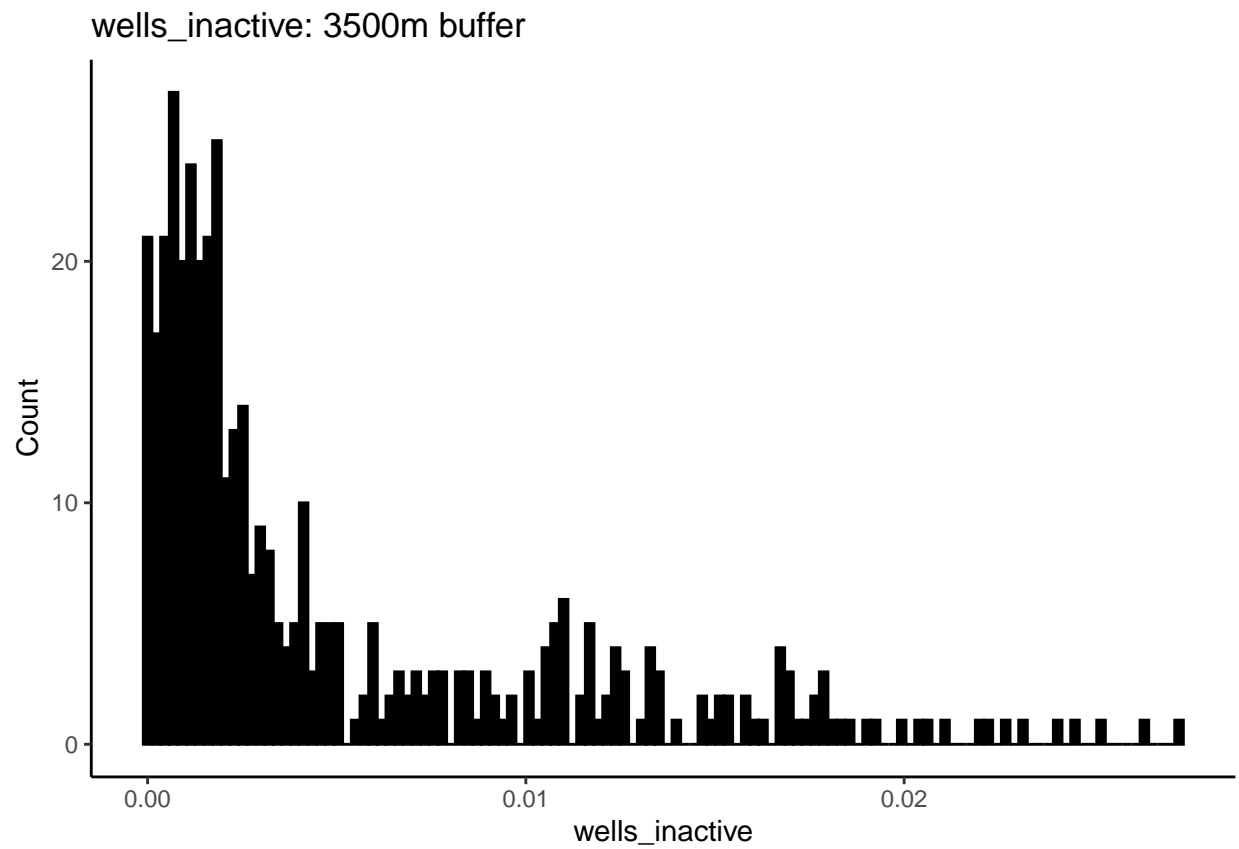
```
##  
## $seismic
```



```
##  
## $wells_active
```

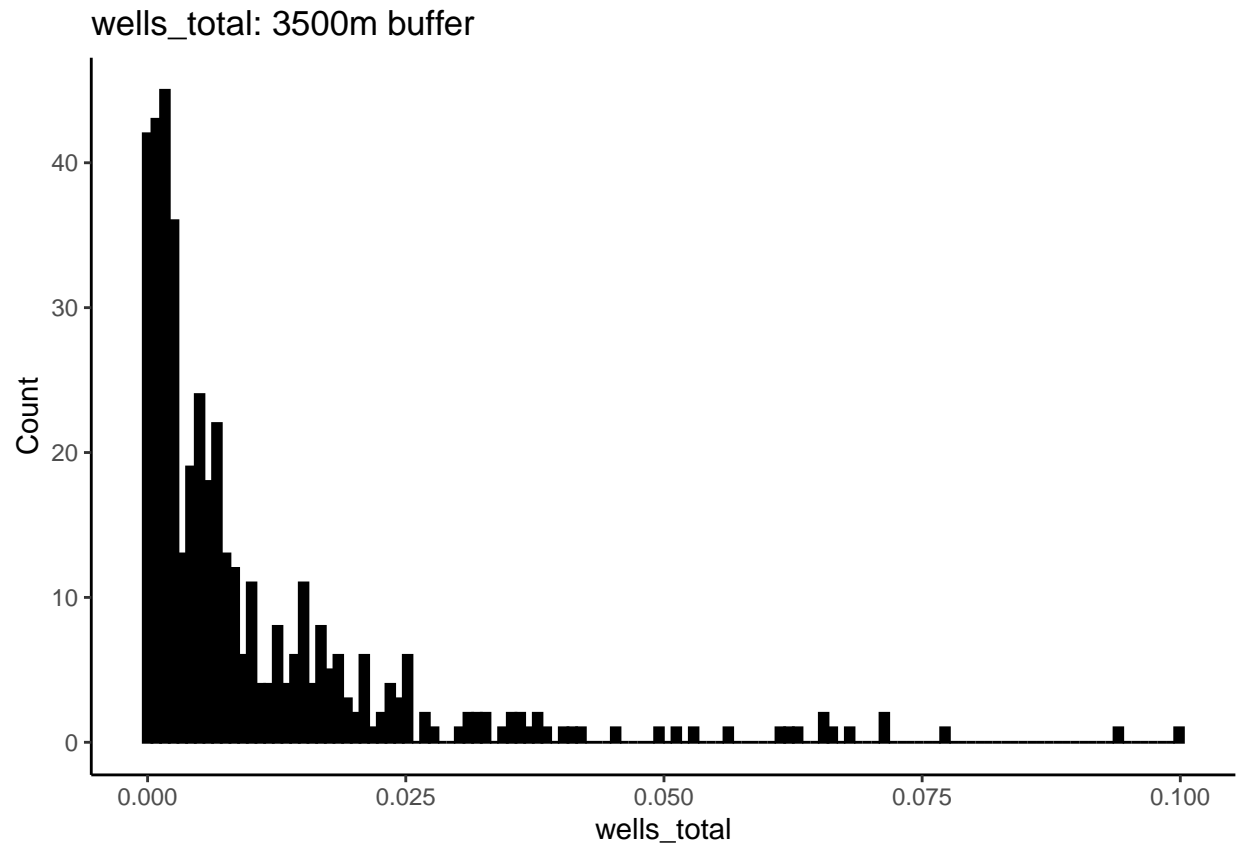


```
##  
## $wells_inactive
```



```
##  
## $wells_total
```



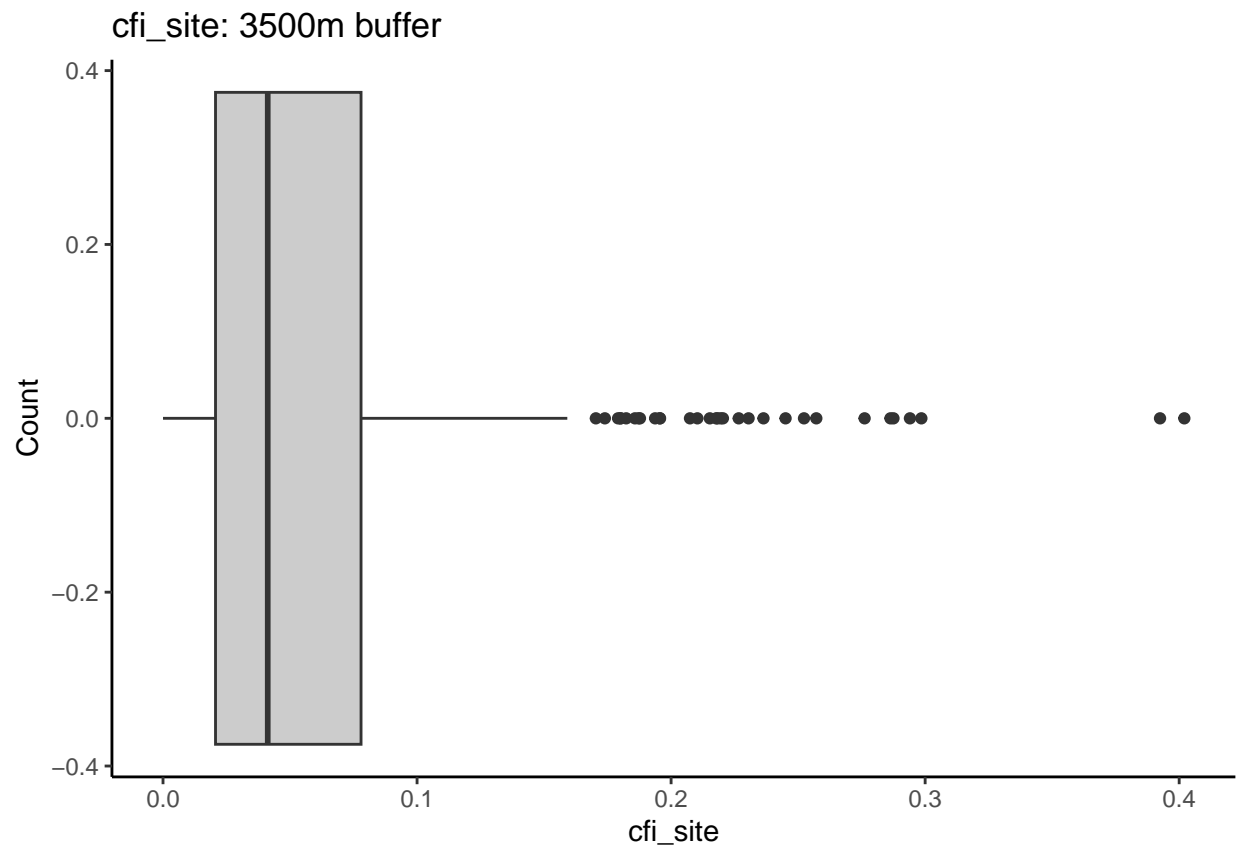


## 2.4. Box and whiskers at 3500 meters:

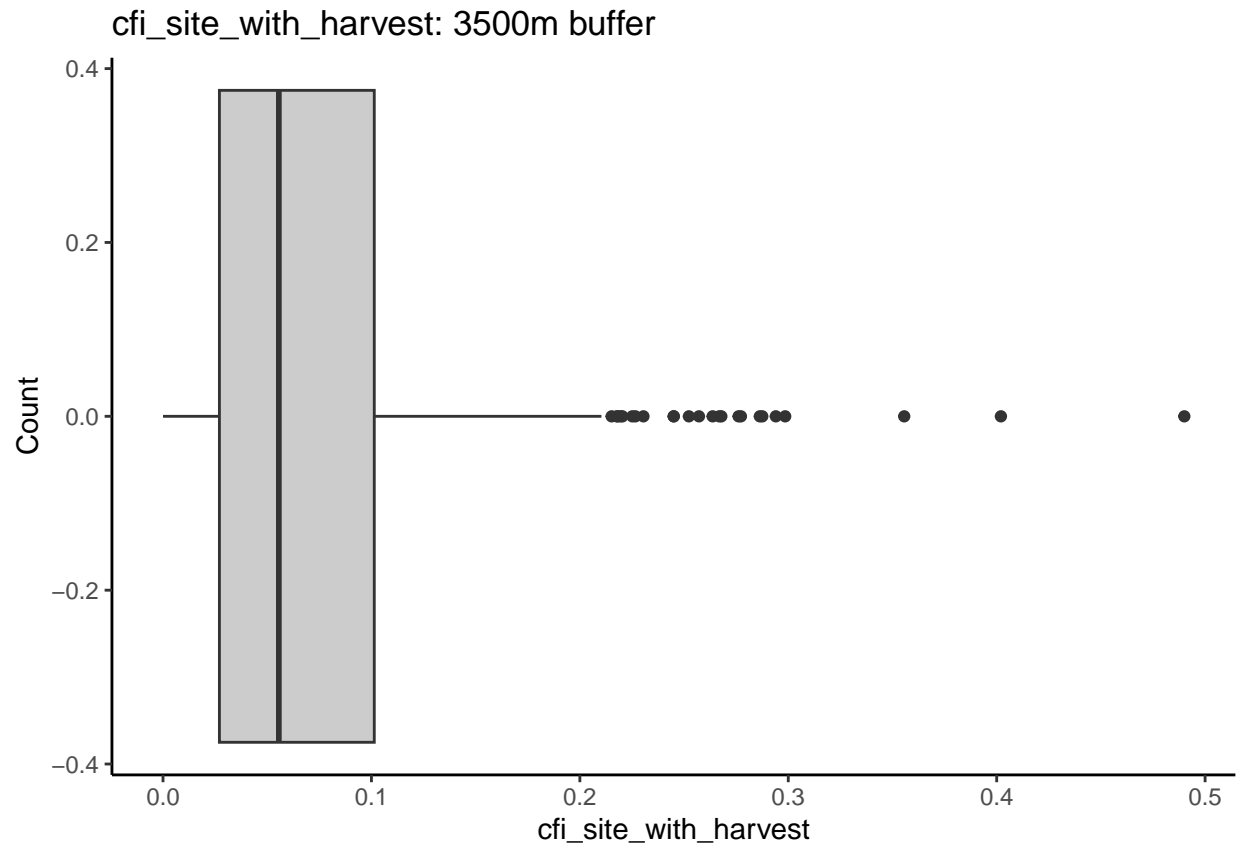
```
covs_to_explore %>%
  filter(buffer_dist == 3500) %>%
  select(-(1:8)) %>%

  # use imap which will retain both the data (x) and the variable names (y)
  imap(~ ggplot(mapping = aes(x = .x)) +
    geom_boxplot(fill = "grey80", color = "grey20") +
    labs(title = paste0(.y, ": 3500m buffer"), x = .y, y = "Count") +
    theme_classic())

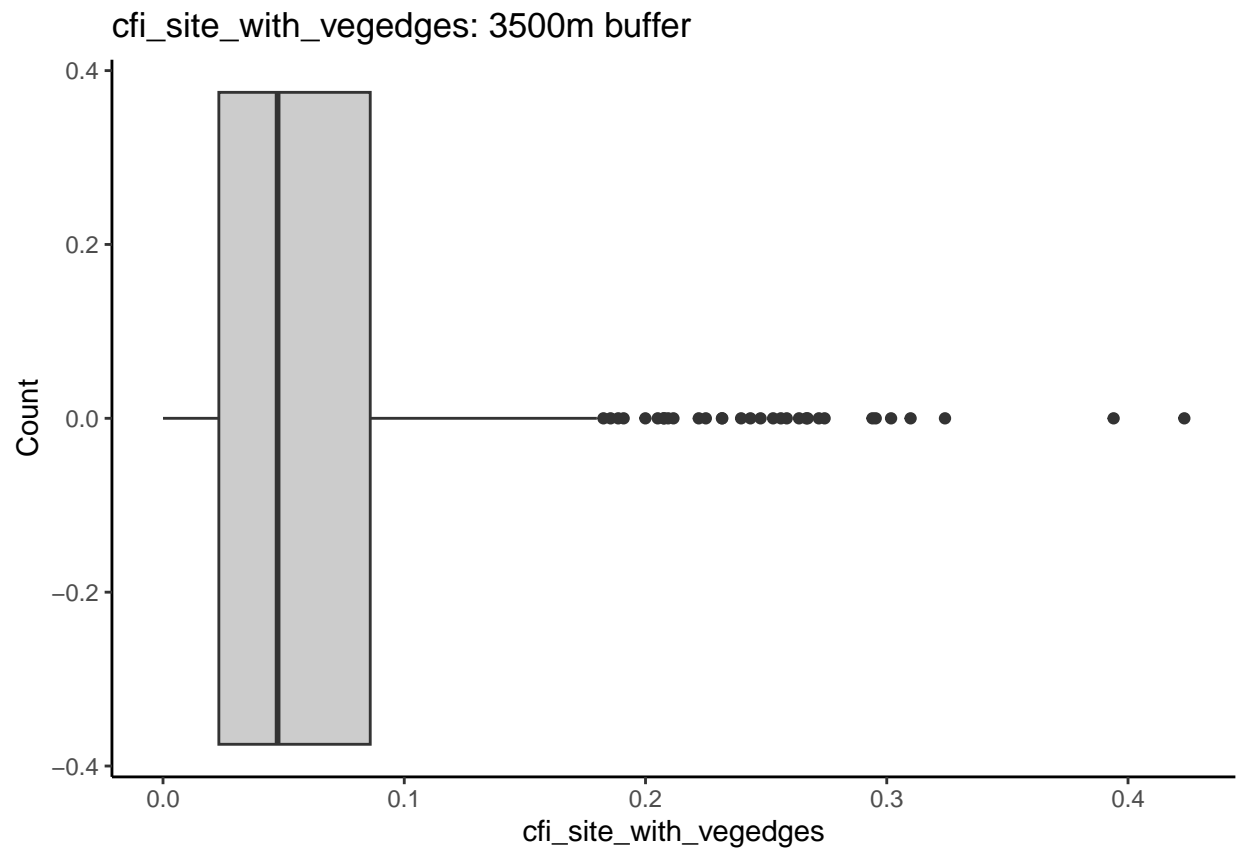
## $cfi_site
```



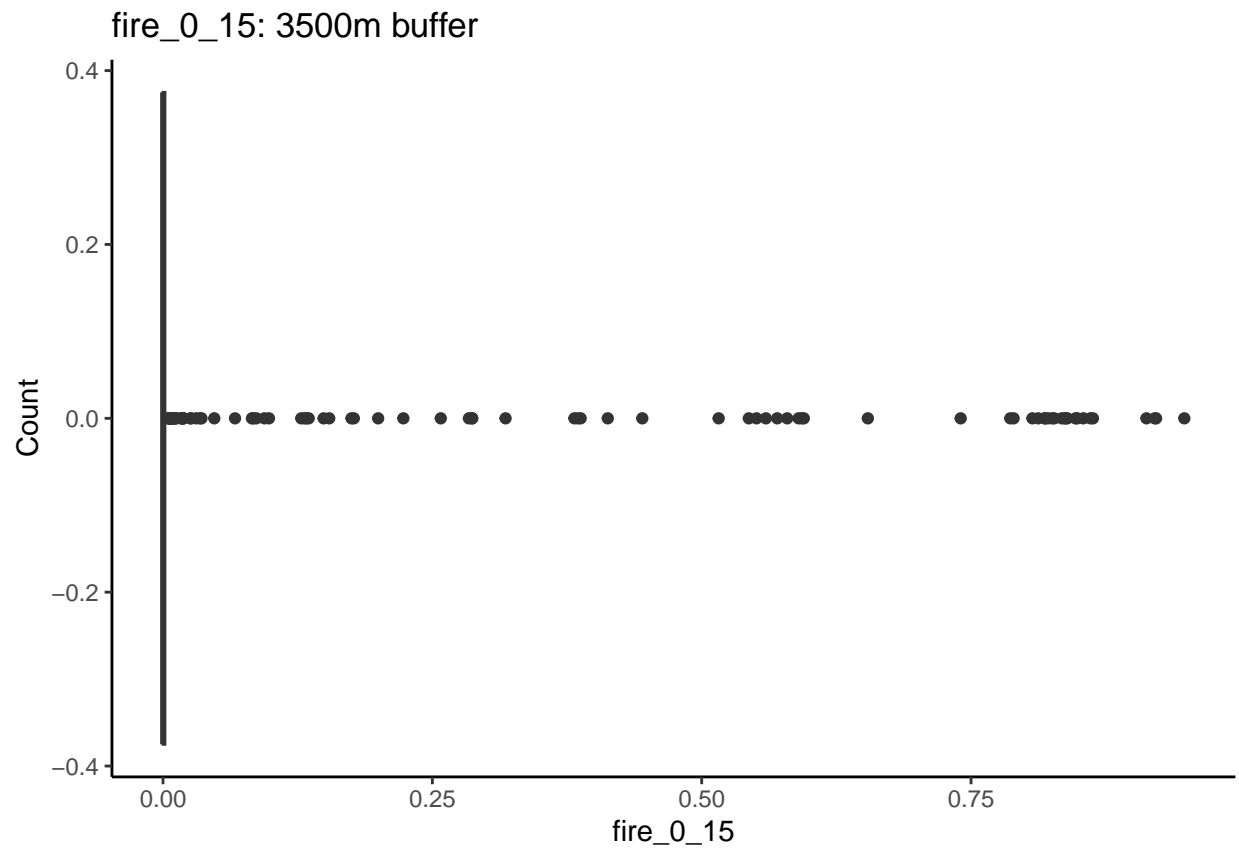
```
##  
## $cfi_site_with_harvest
```



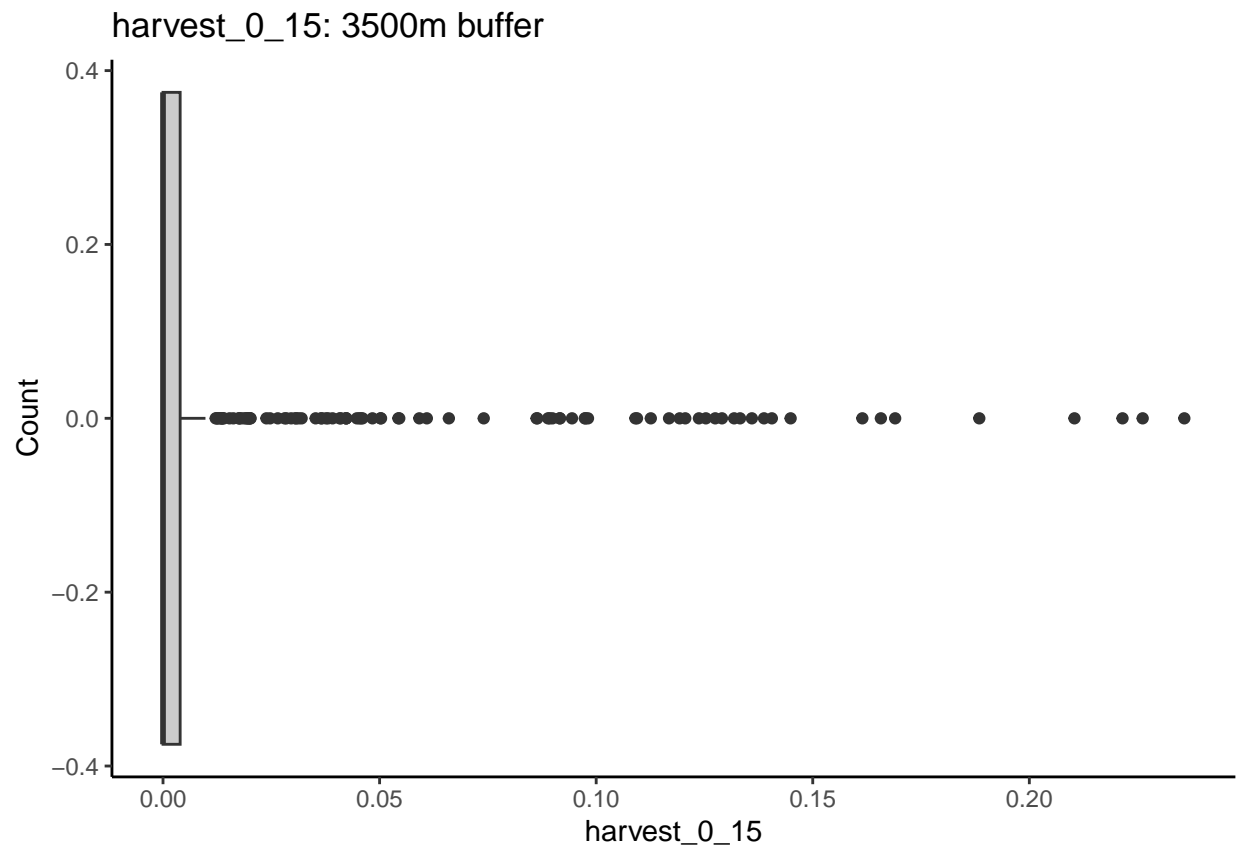
```
##  
## $cfi_site_with_vegedges
```



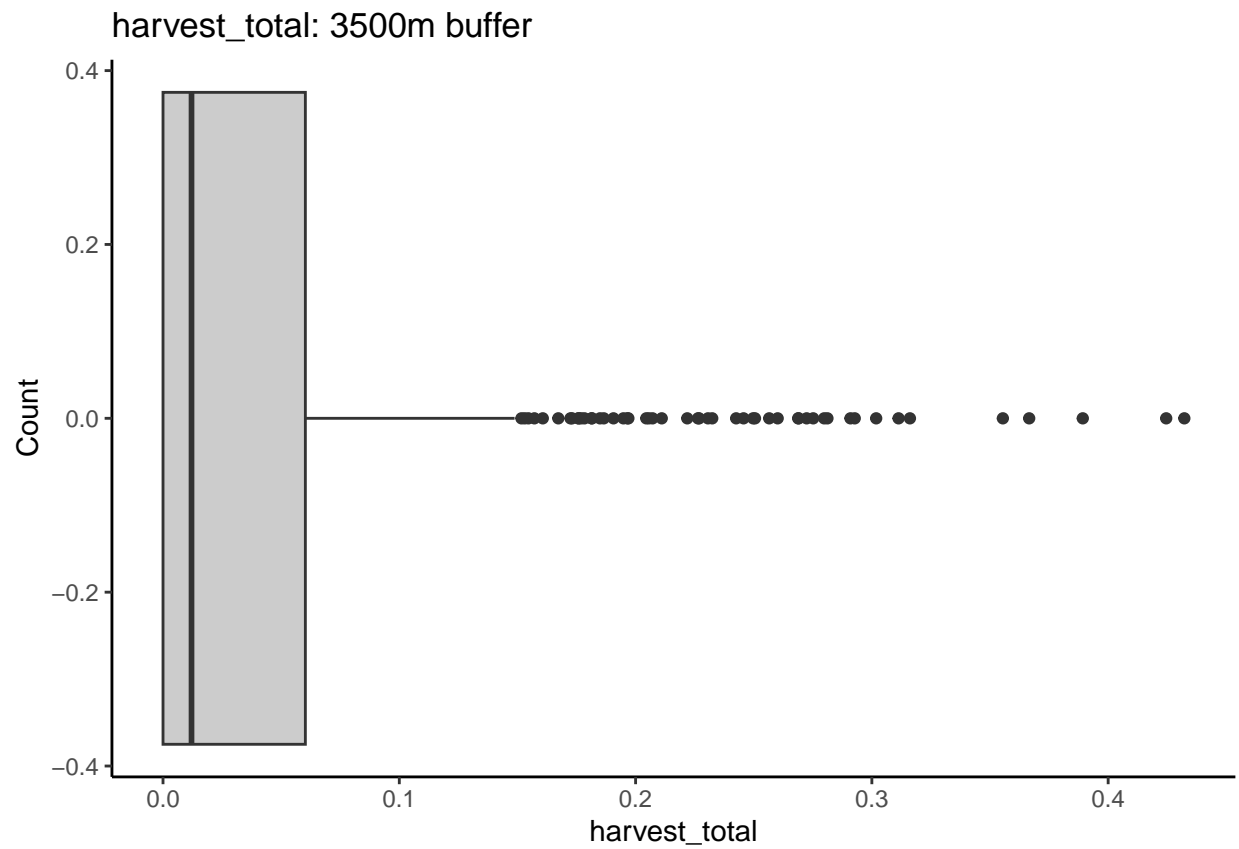
```
##  
## $fire_0_15
```



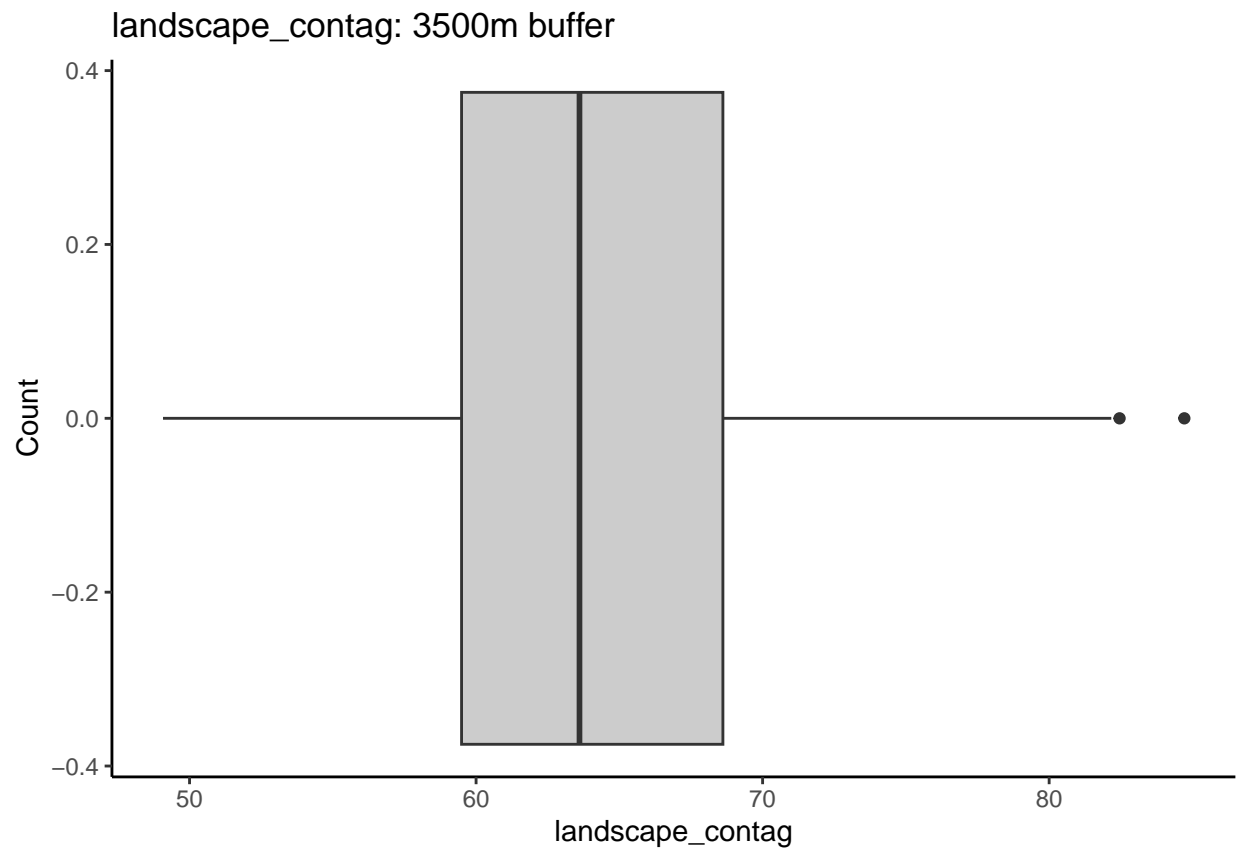
##  
## \$harvest\_0\_15



```
##  
## $harvest_total
```

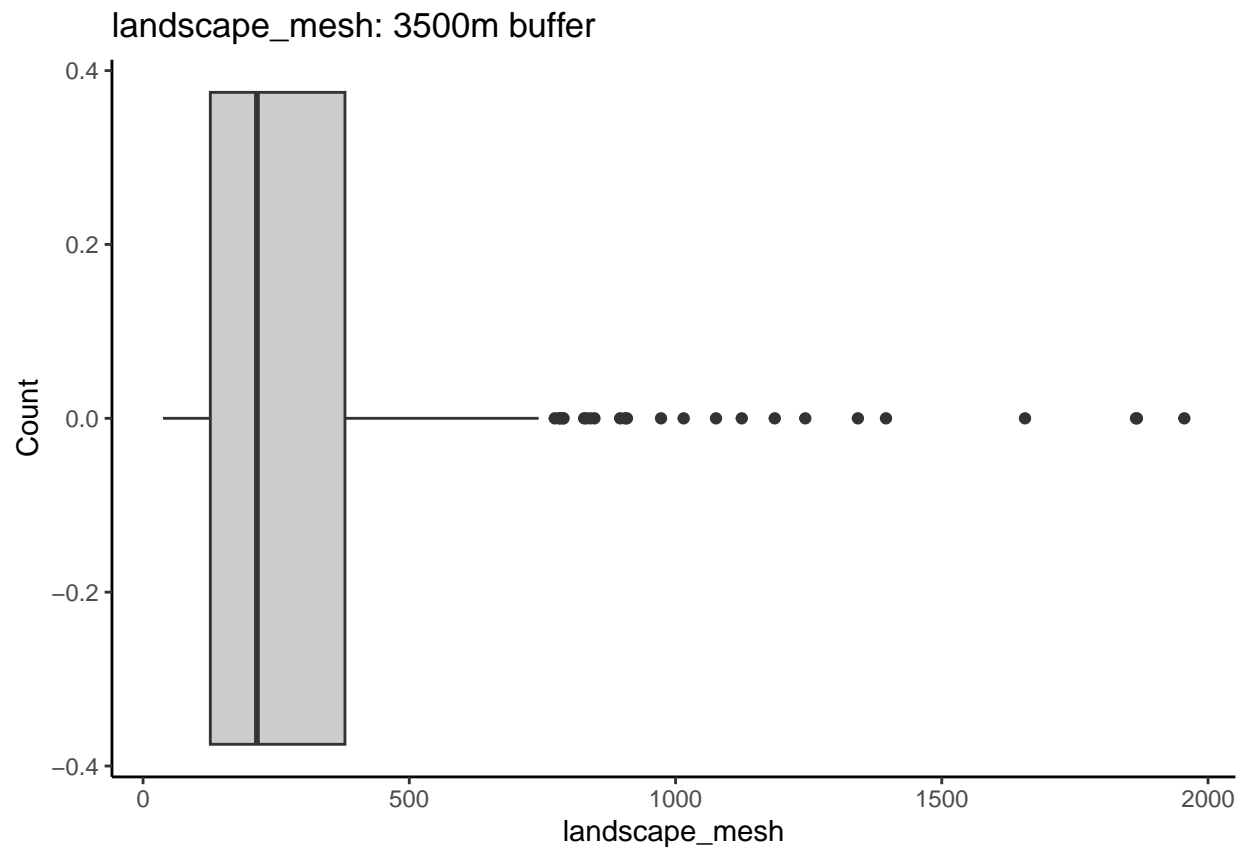


```
##  
## $landscape_contag
```

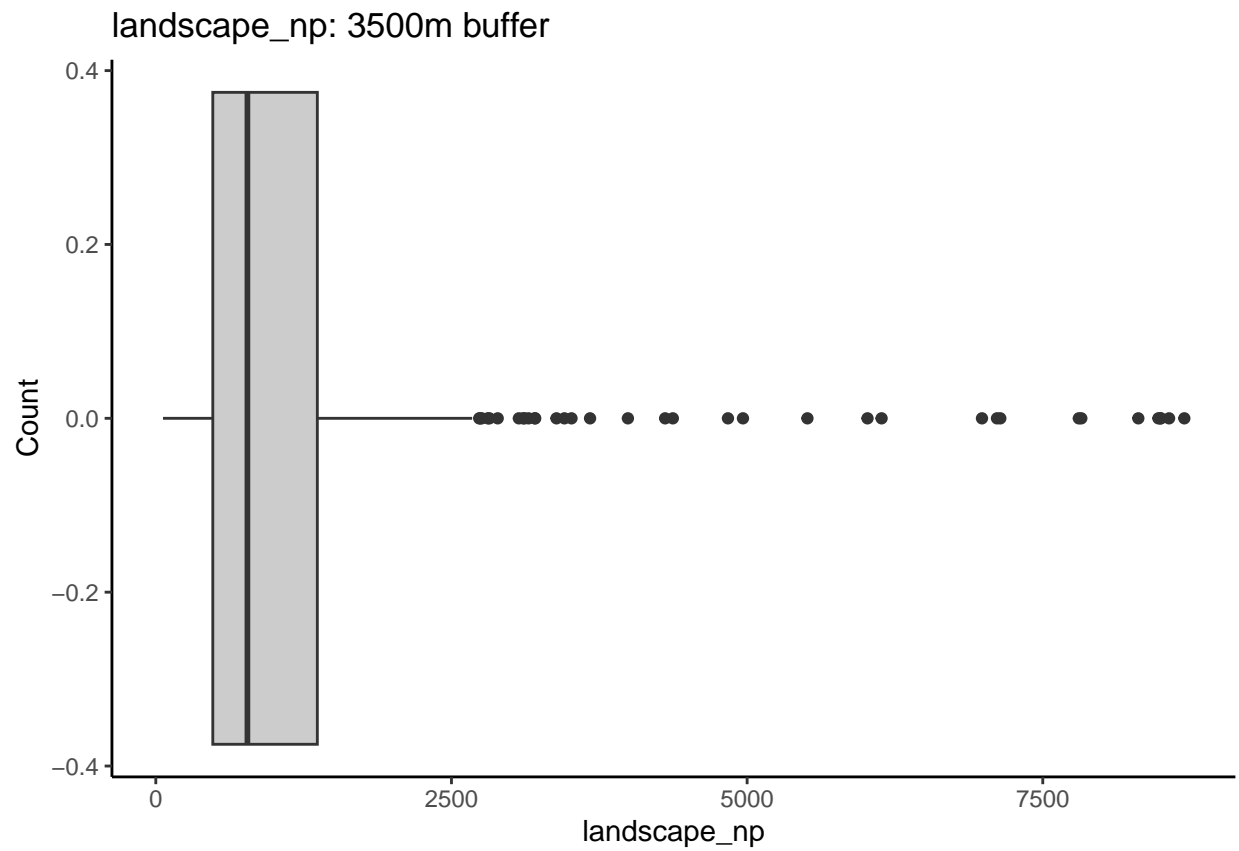


```
##  
## $landscape_mesh
```

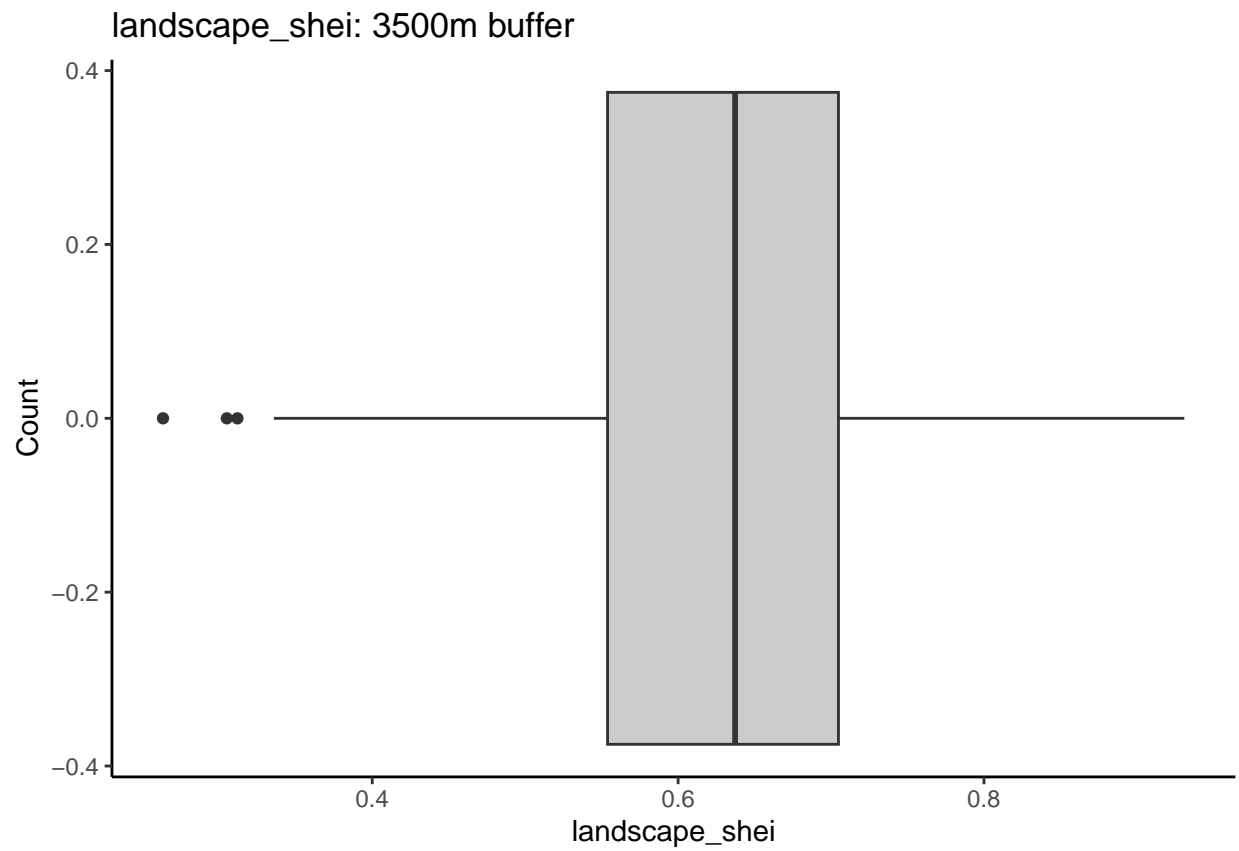




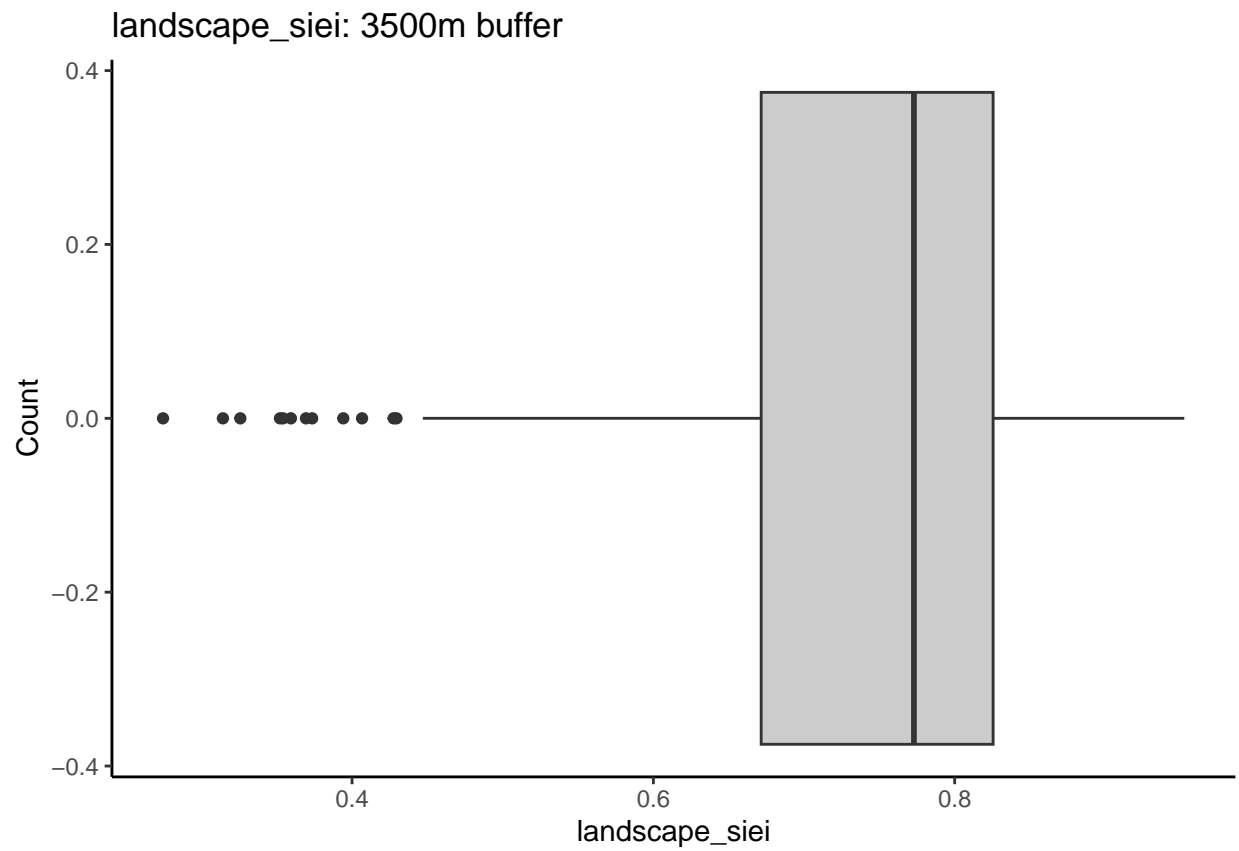
```
##  
## $landscape_np
```



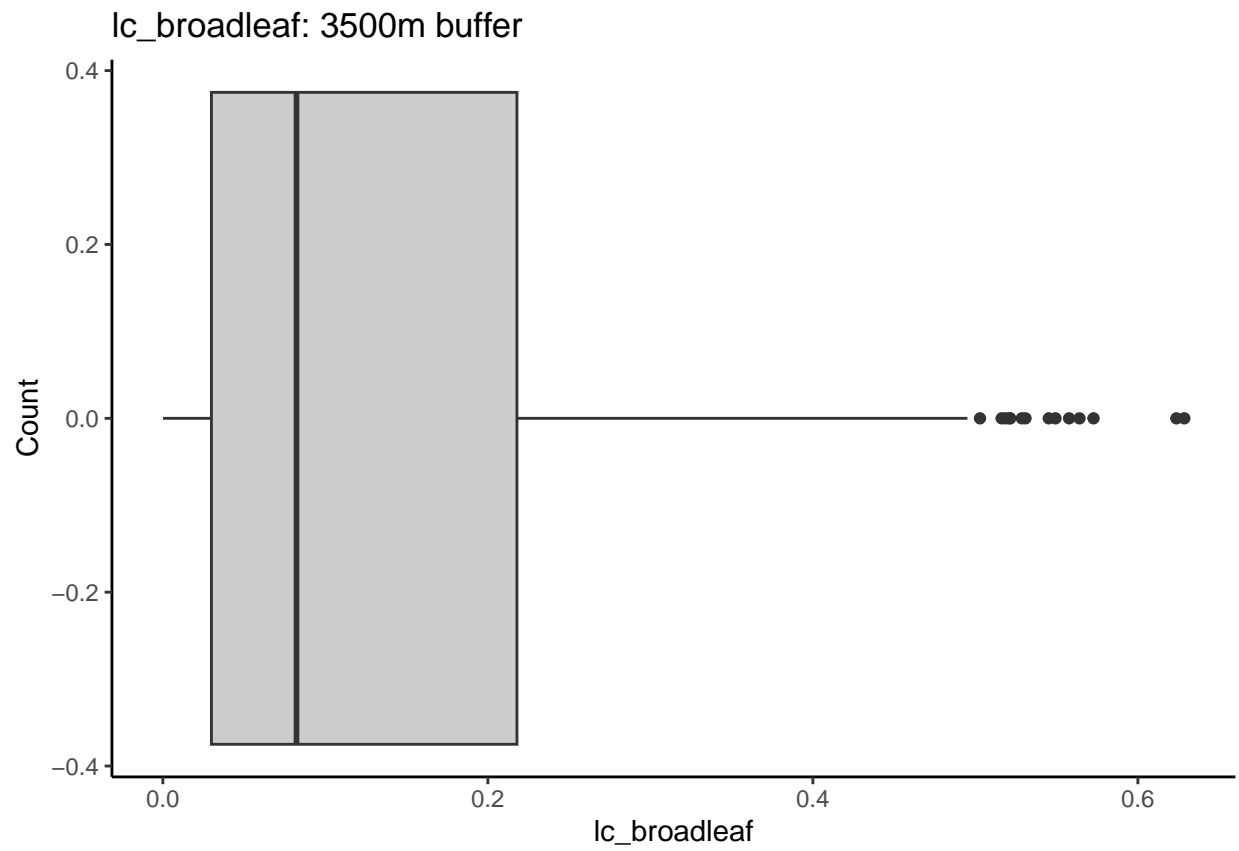
```
##  
## $landscape_shei
```



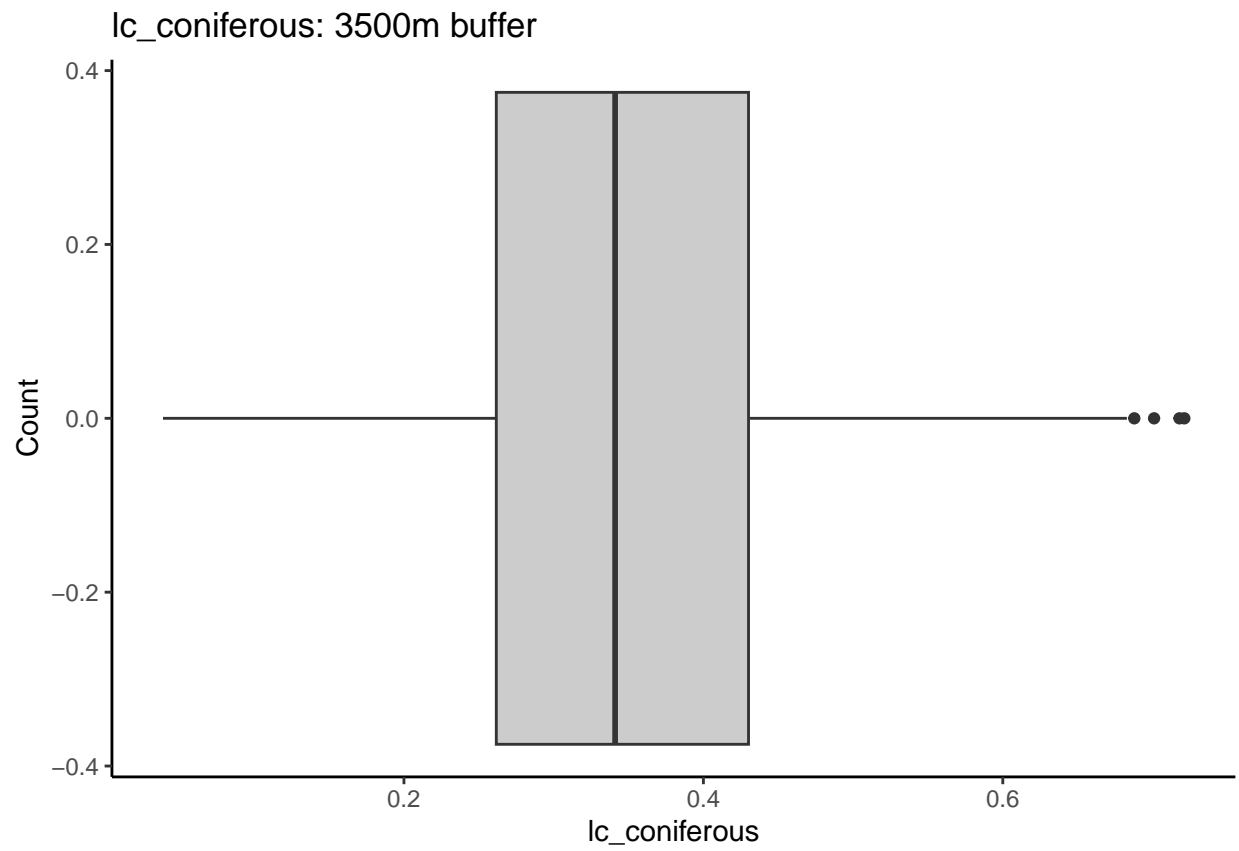
```
##  
## $landscape_siei
```



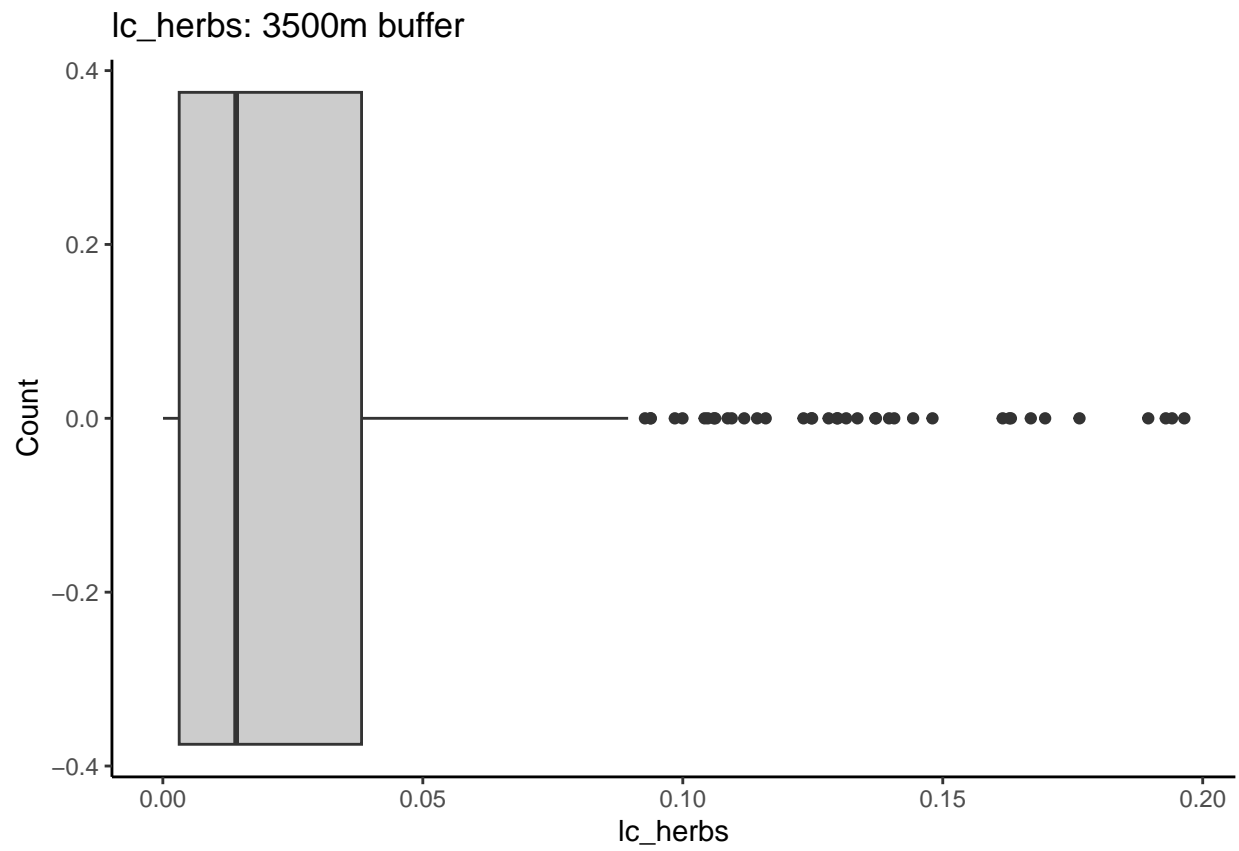
```
##  
## $lc_broadleaf
```



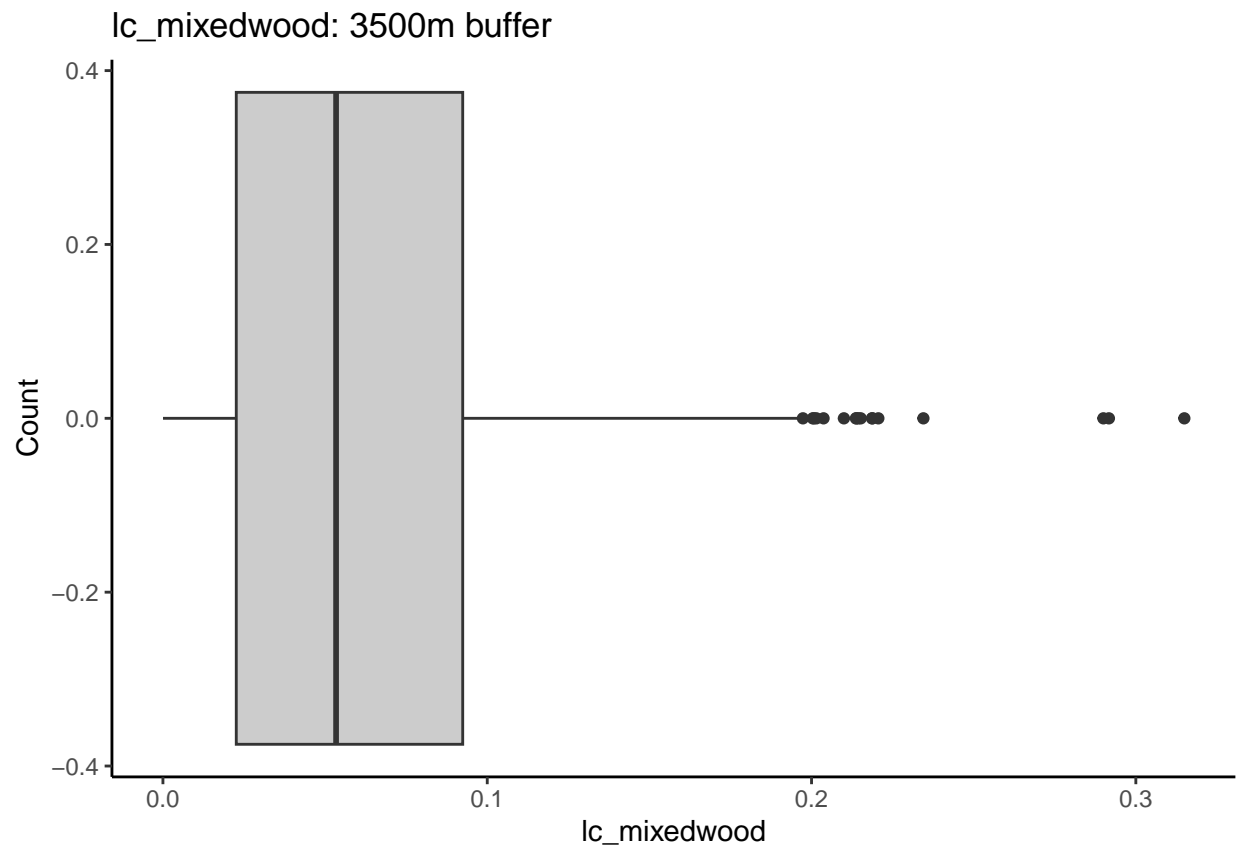
```
##  
## $lc_coniferous
```



```
##  
## $lc_herbs
```

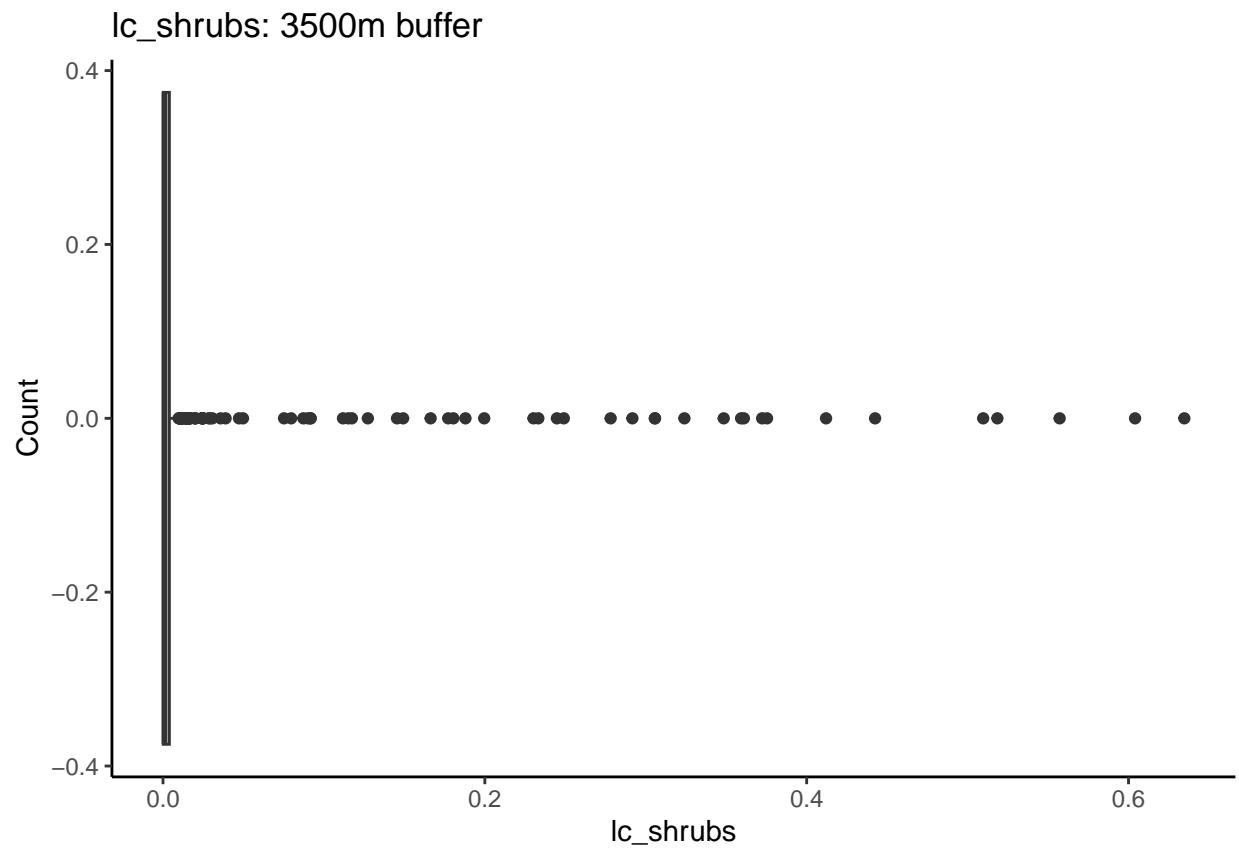


```
##  
## $lc_mixedwood
```

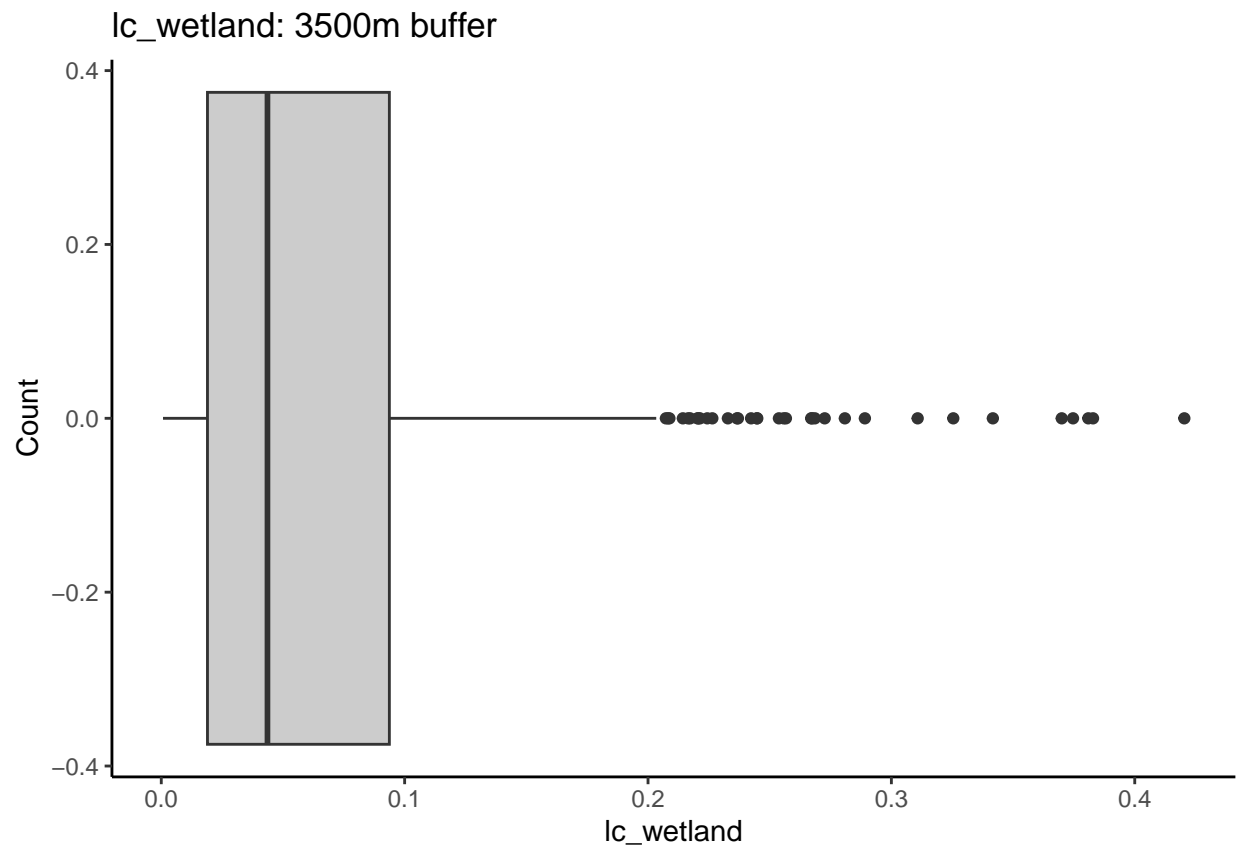


```
##  
## $lc_shrubs
```

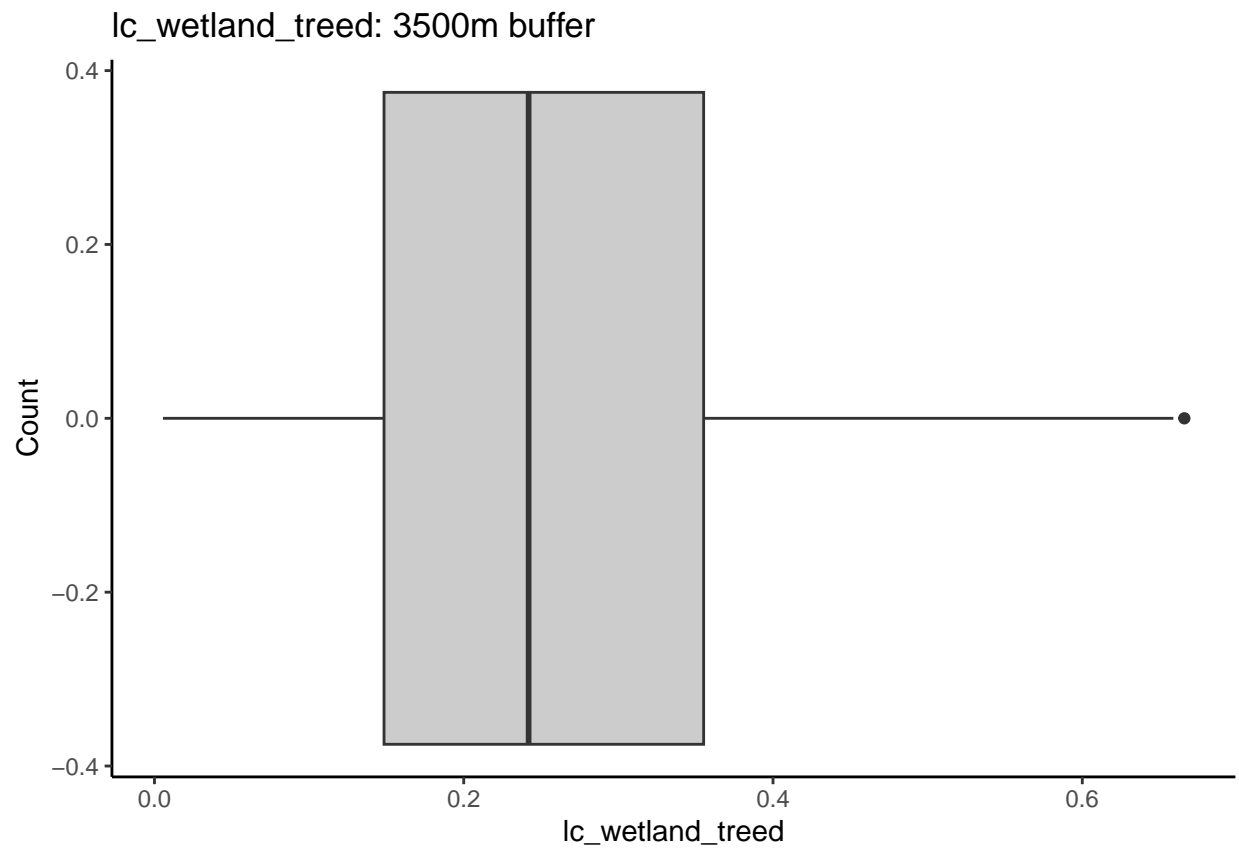




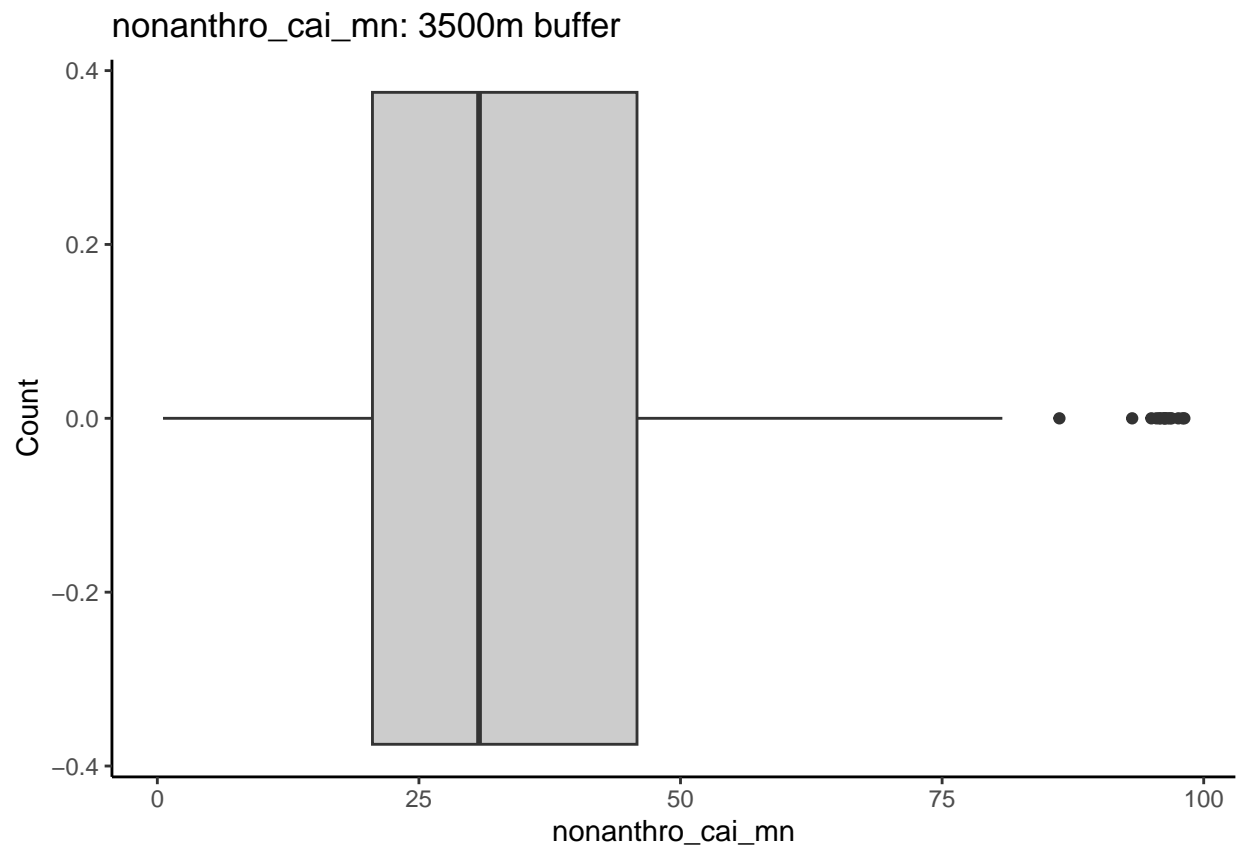
```
##  
## $lc_wetland
```



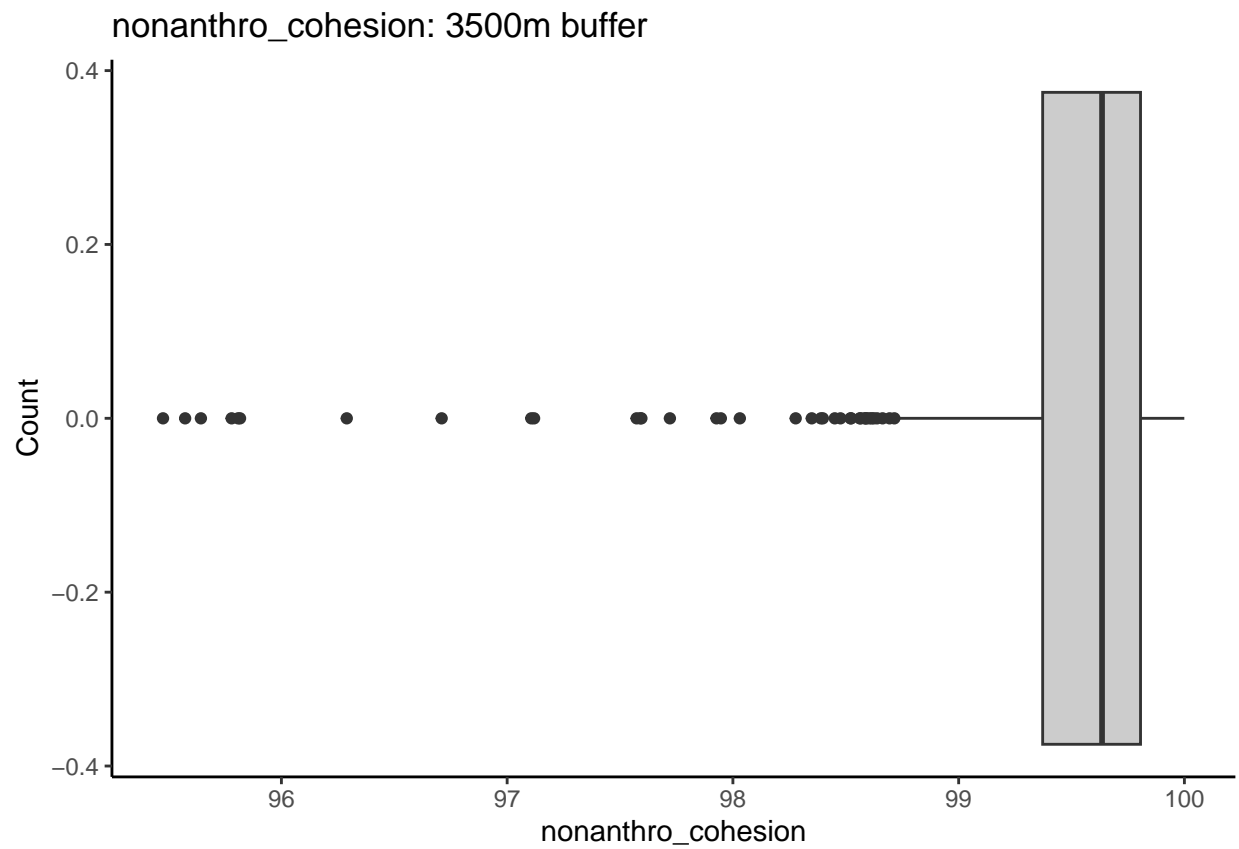
```
##  
## $lc_wetland_treed
```



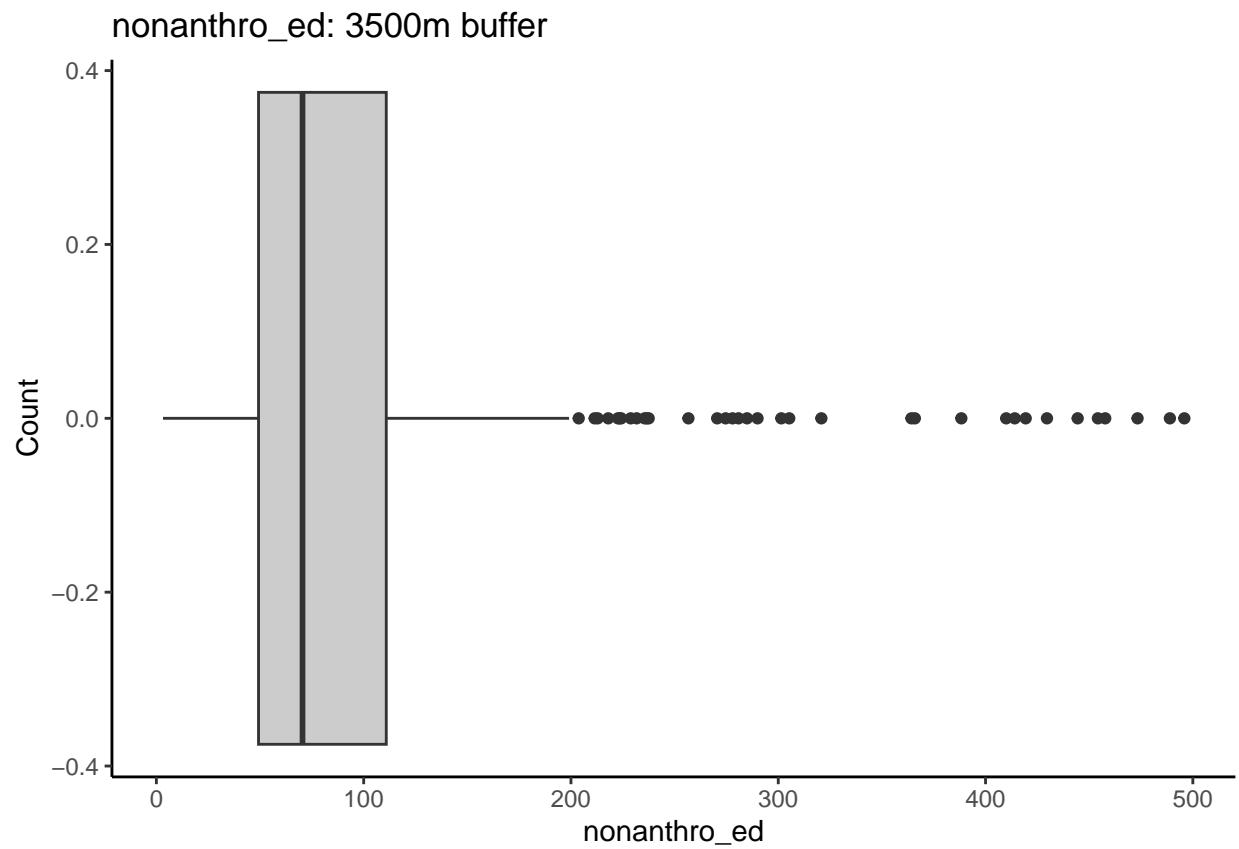
```
##  
## $nonanthro_cai_mn
```



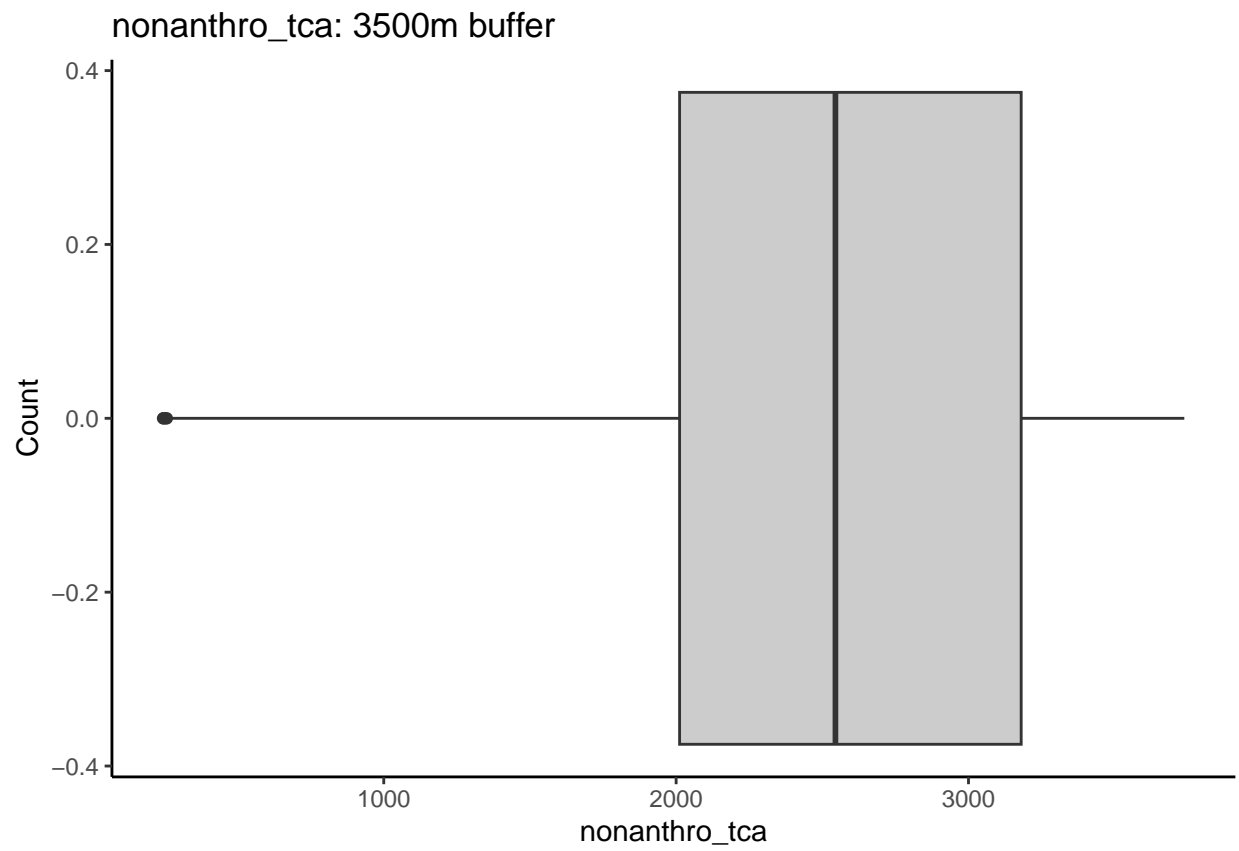
```
##  
## $nonanthro_cohesion
```



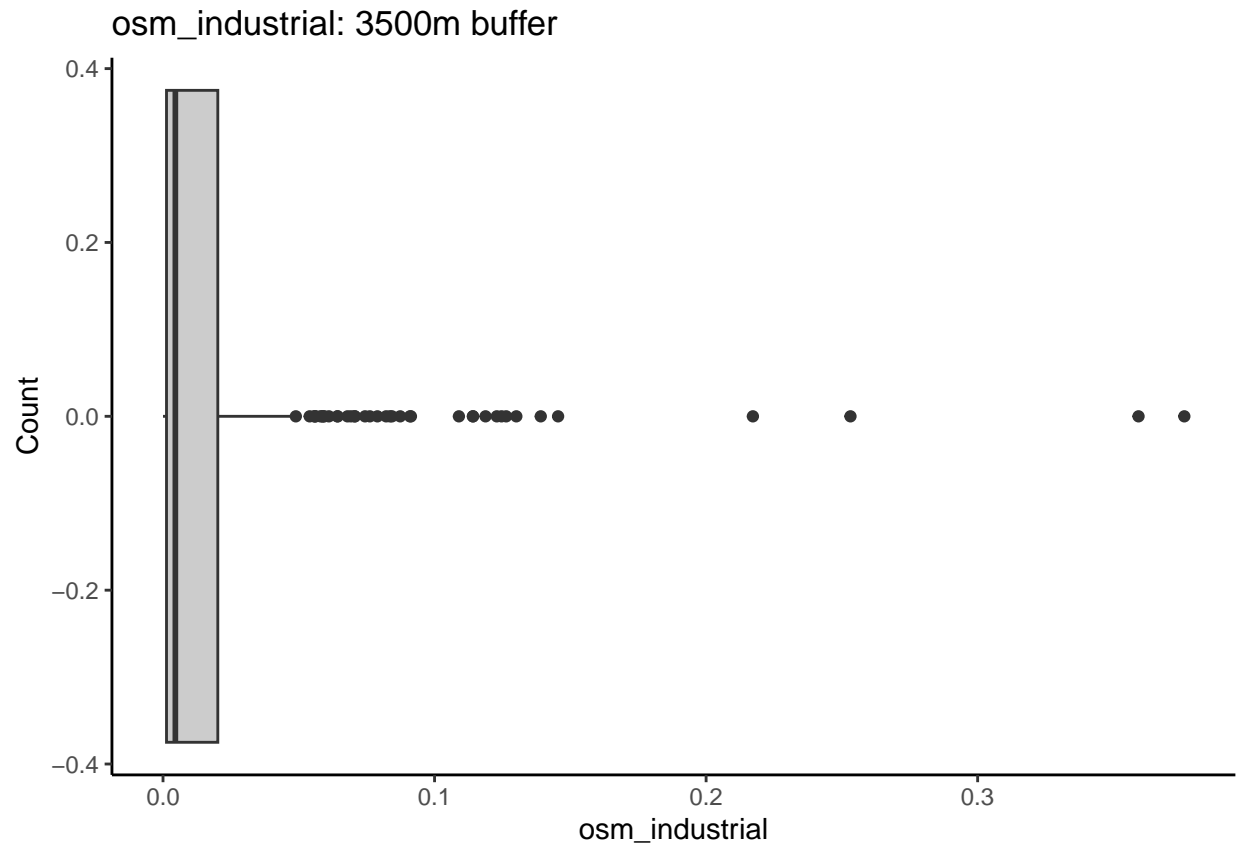
```
##  
## $nonanthro_ed
```



```
##  
## $nonanthro_tca
```

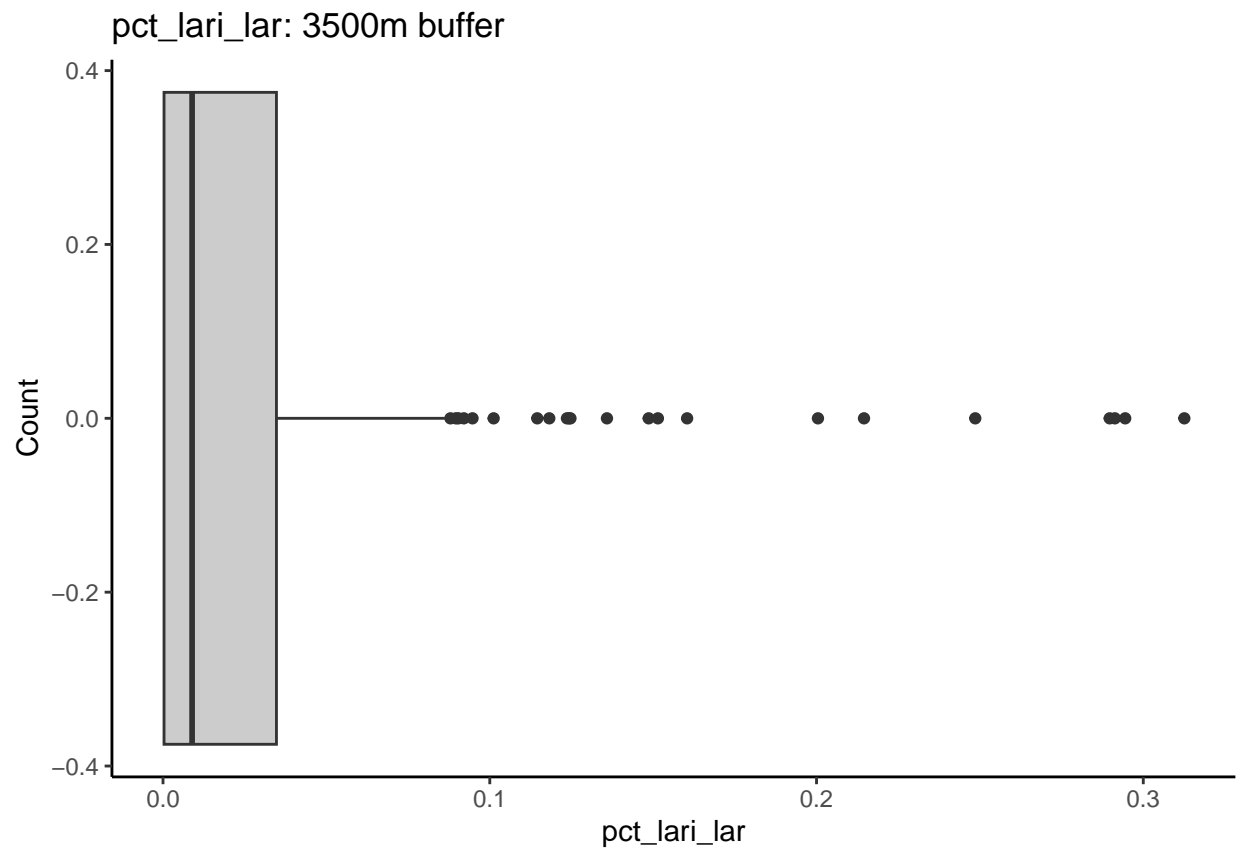


```
##  
## $osm_industrial
```

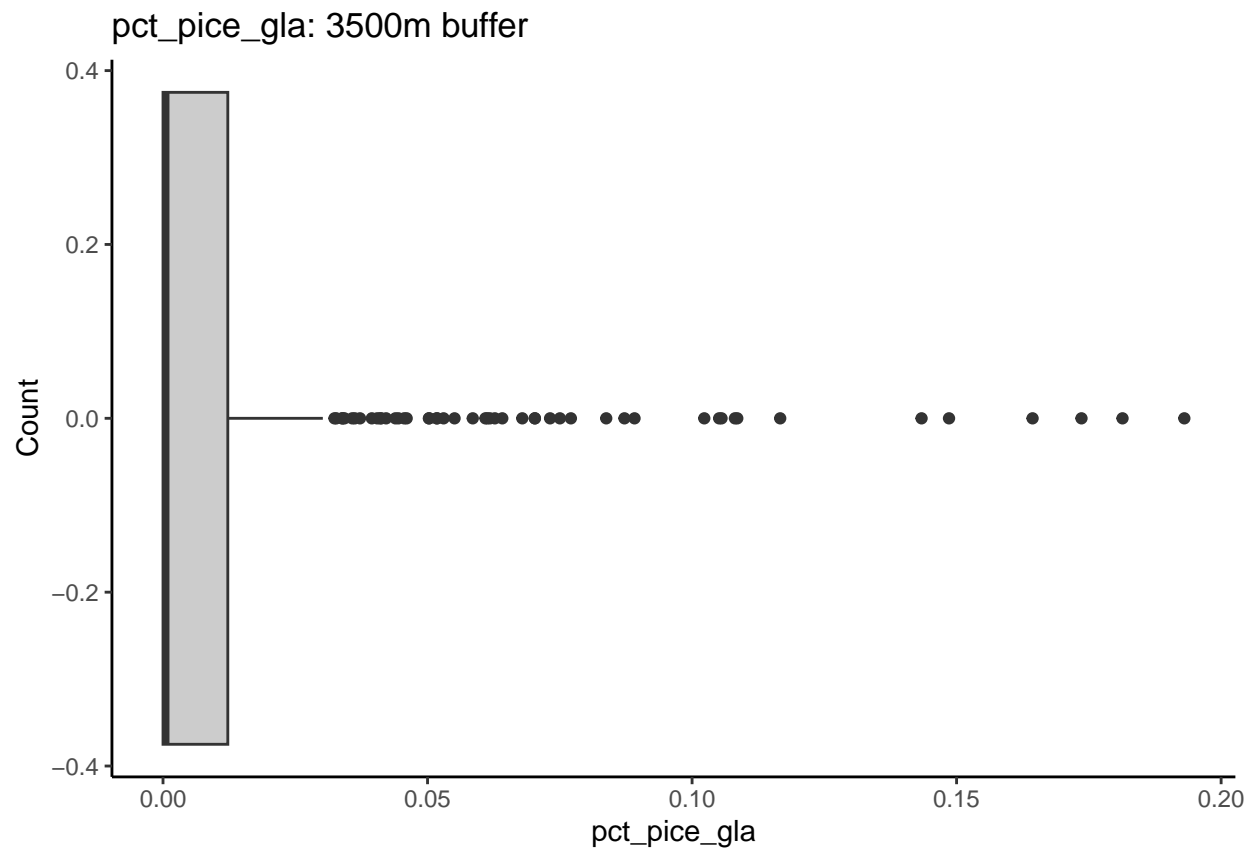


```
##  
## $pct_lari_lar
```

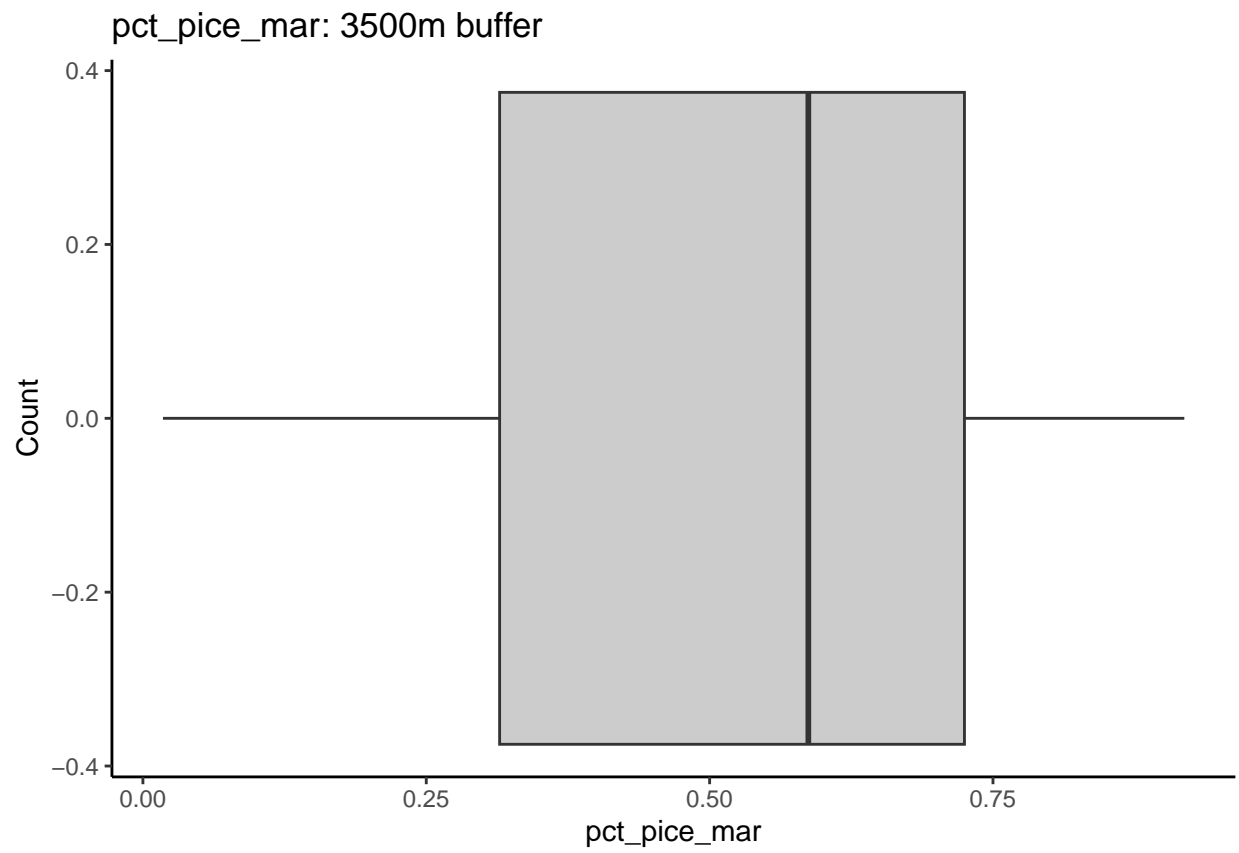




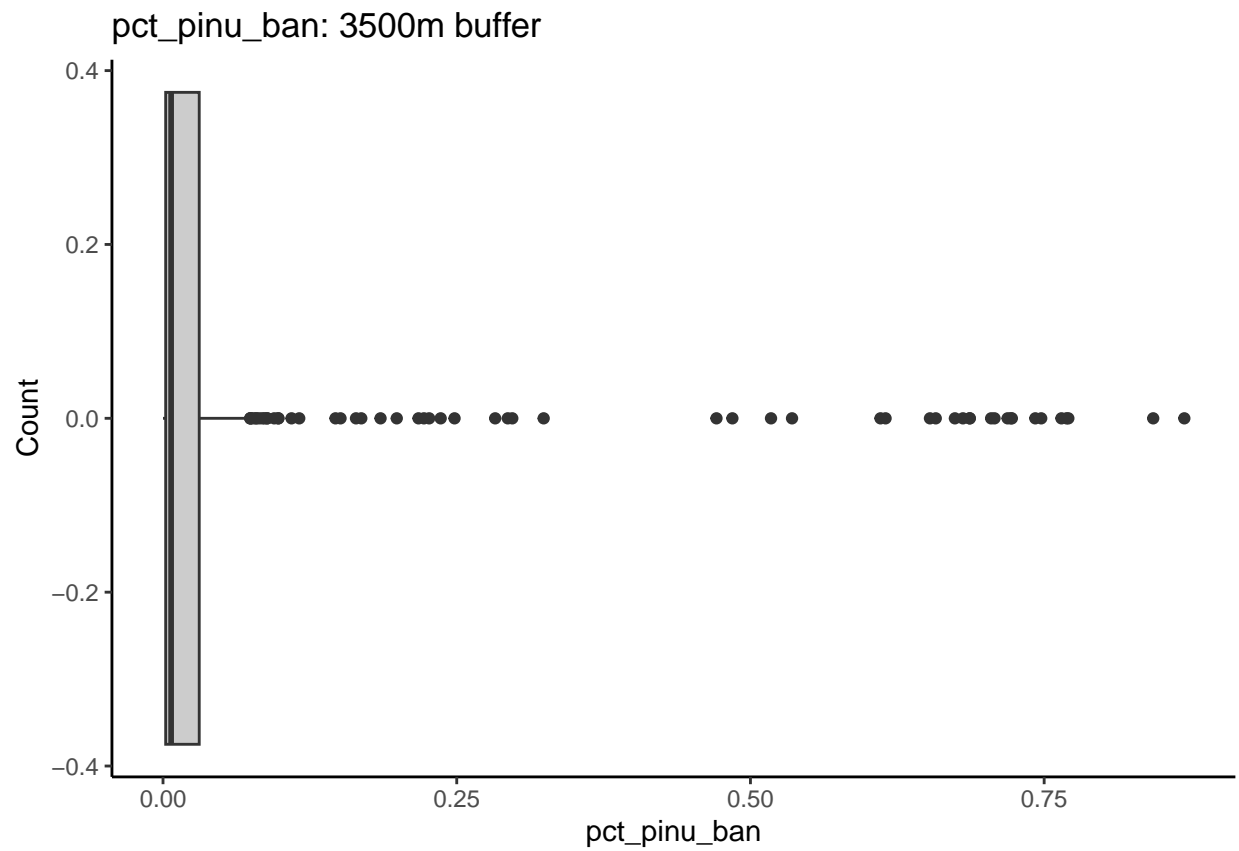
```
##  
## $pct_pice_gla
```



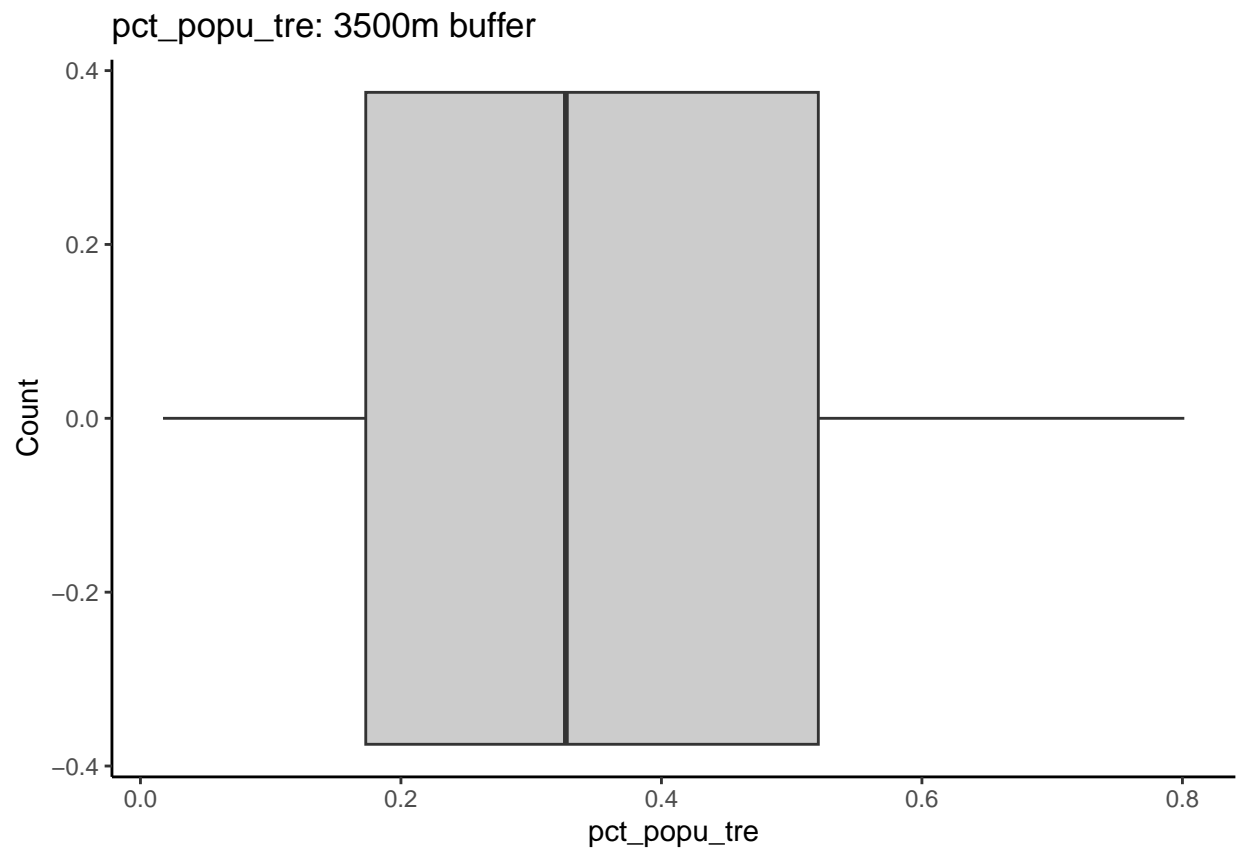
```
##  
## $pct_pice_mar
```



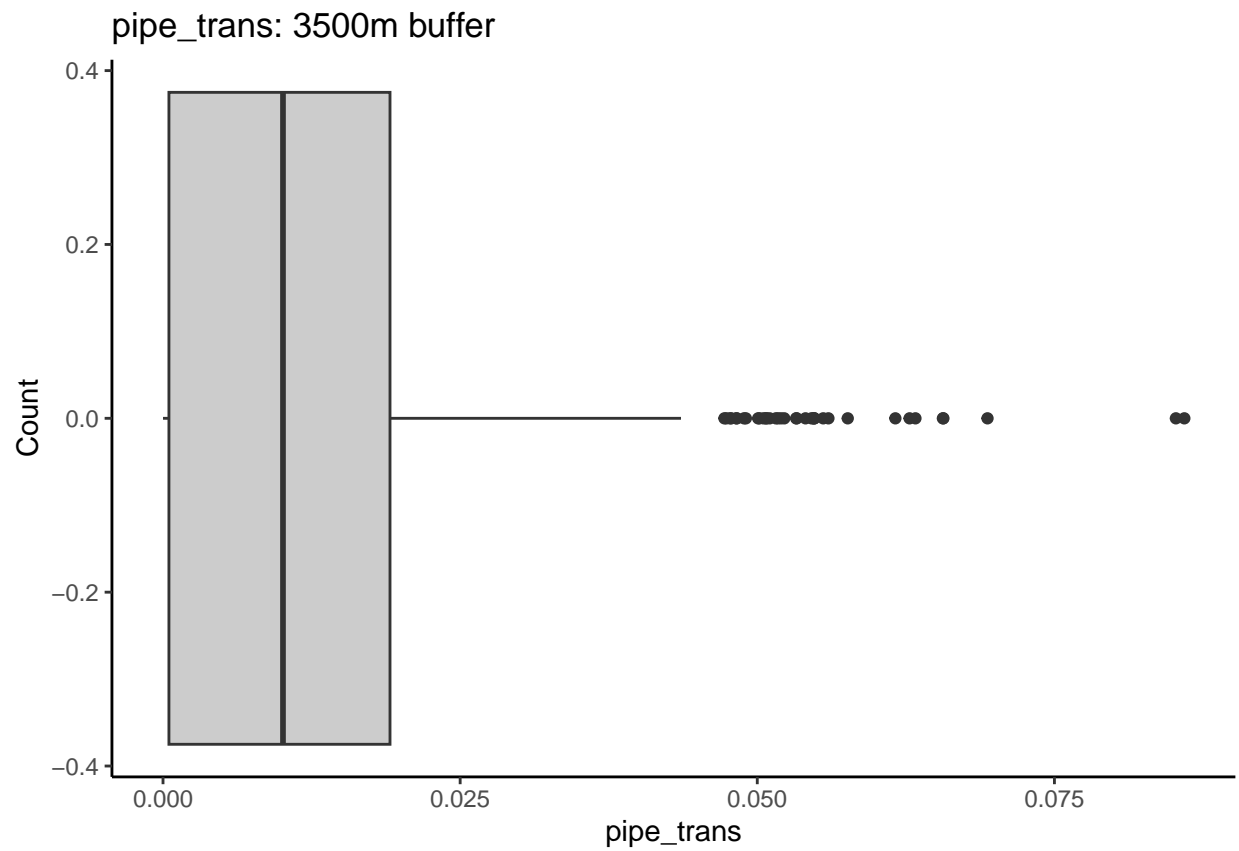
```
##  
## $pct_pinu_ban
```



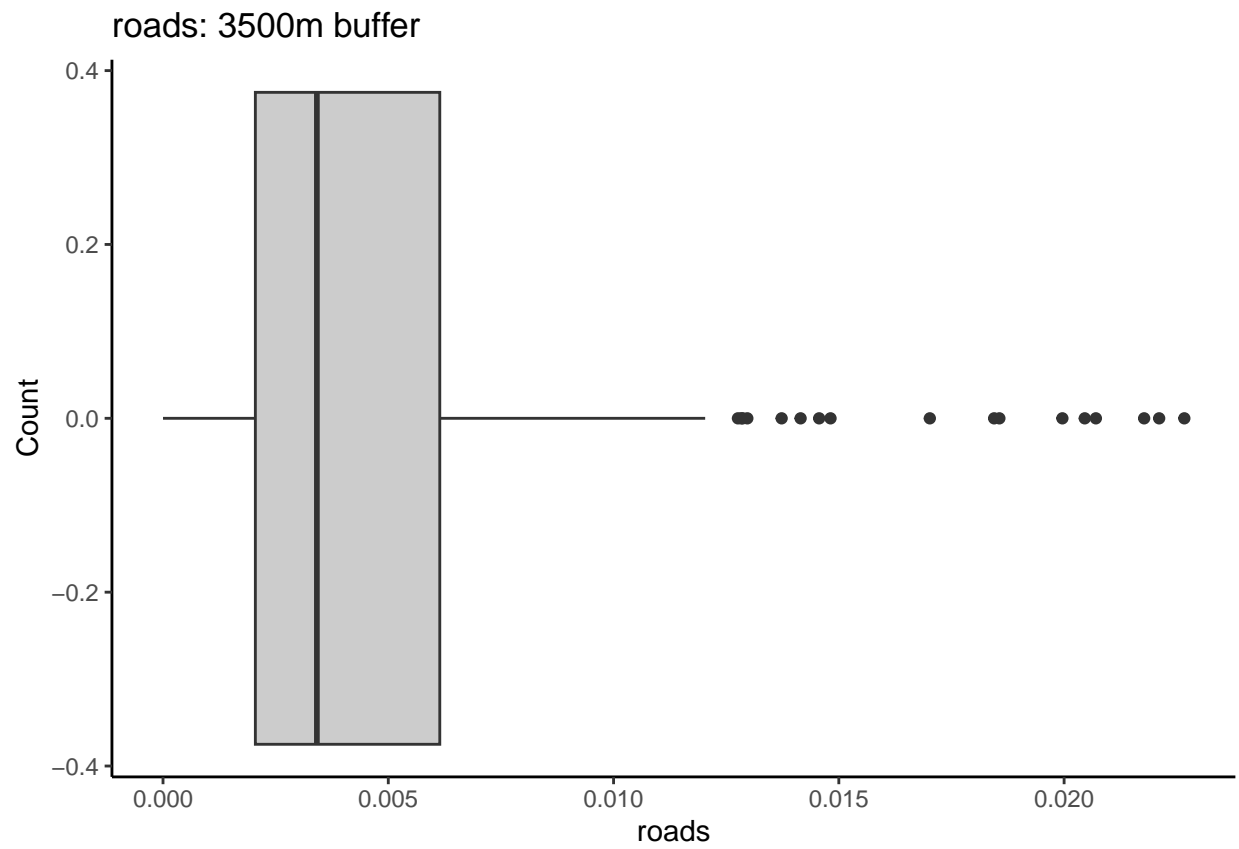
```
##  
## $pct_popu_tre
```



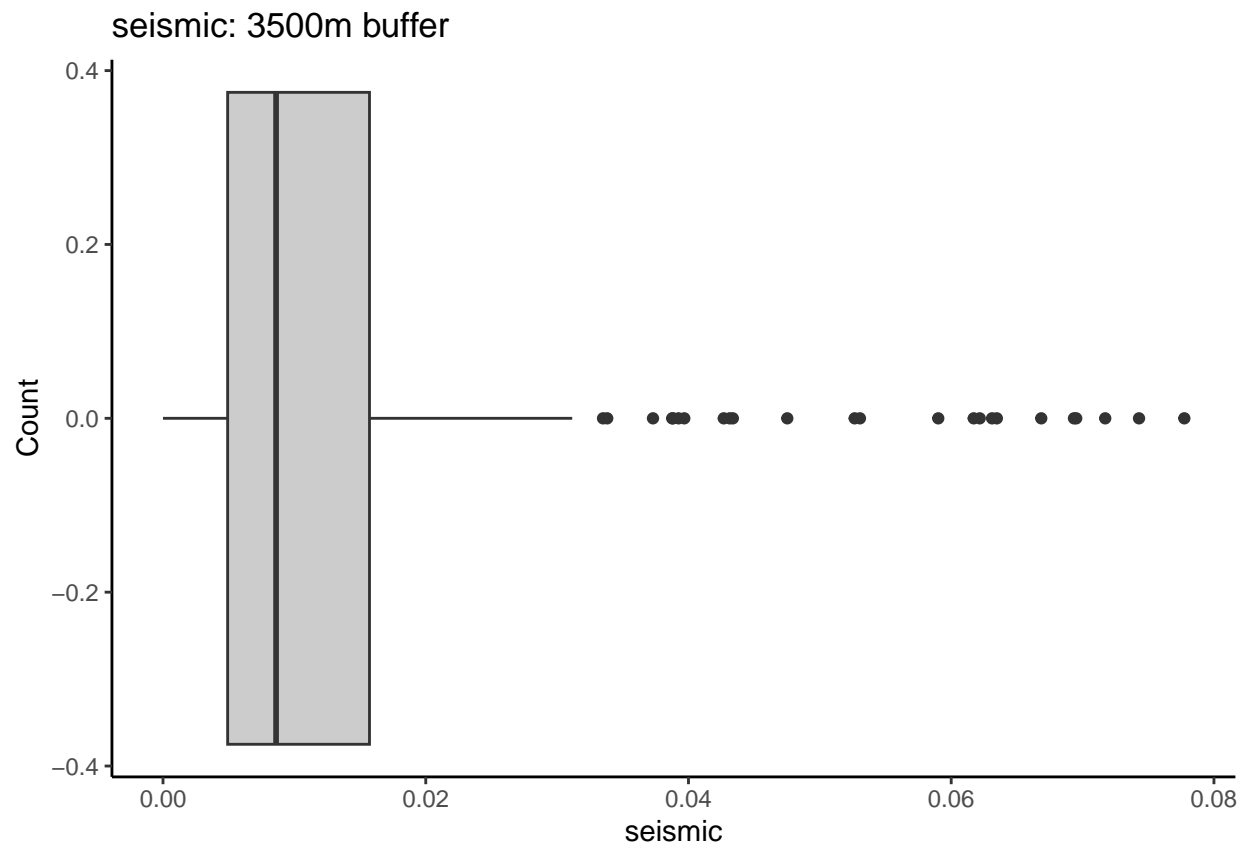
```
##  
## $pipe_trans
```



##  
## \$roads

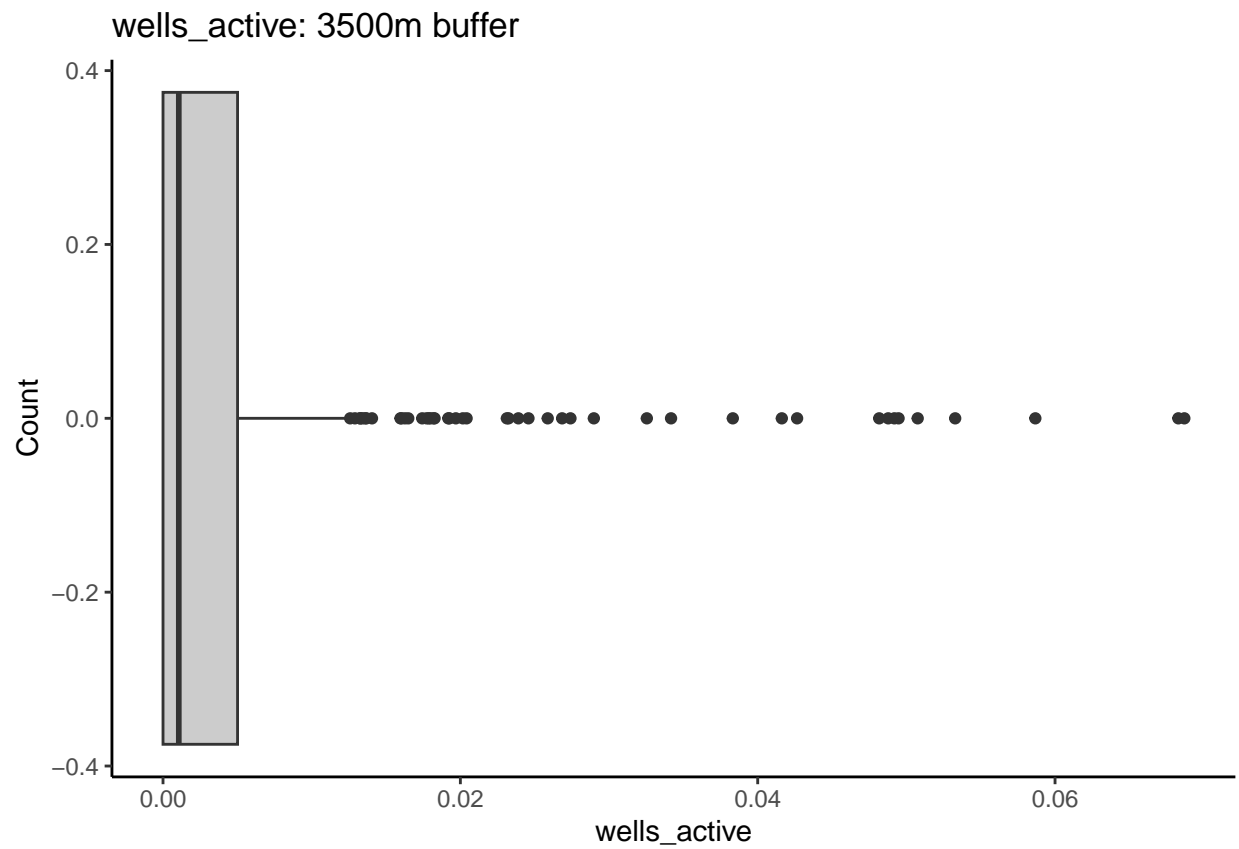


```
##  
## $seismic
```

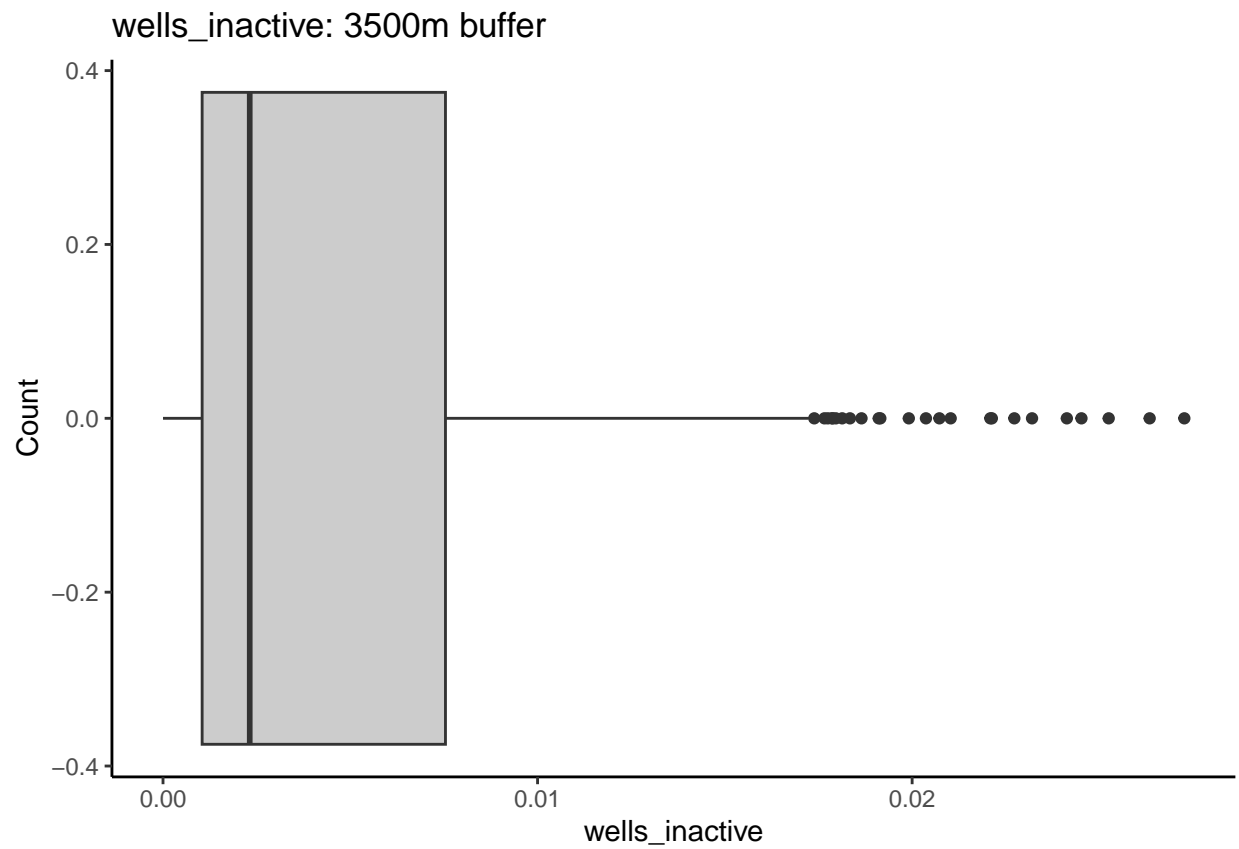


```
##  
## $wells_active
```

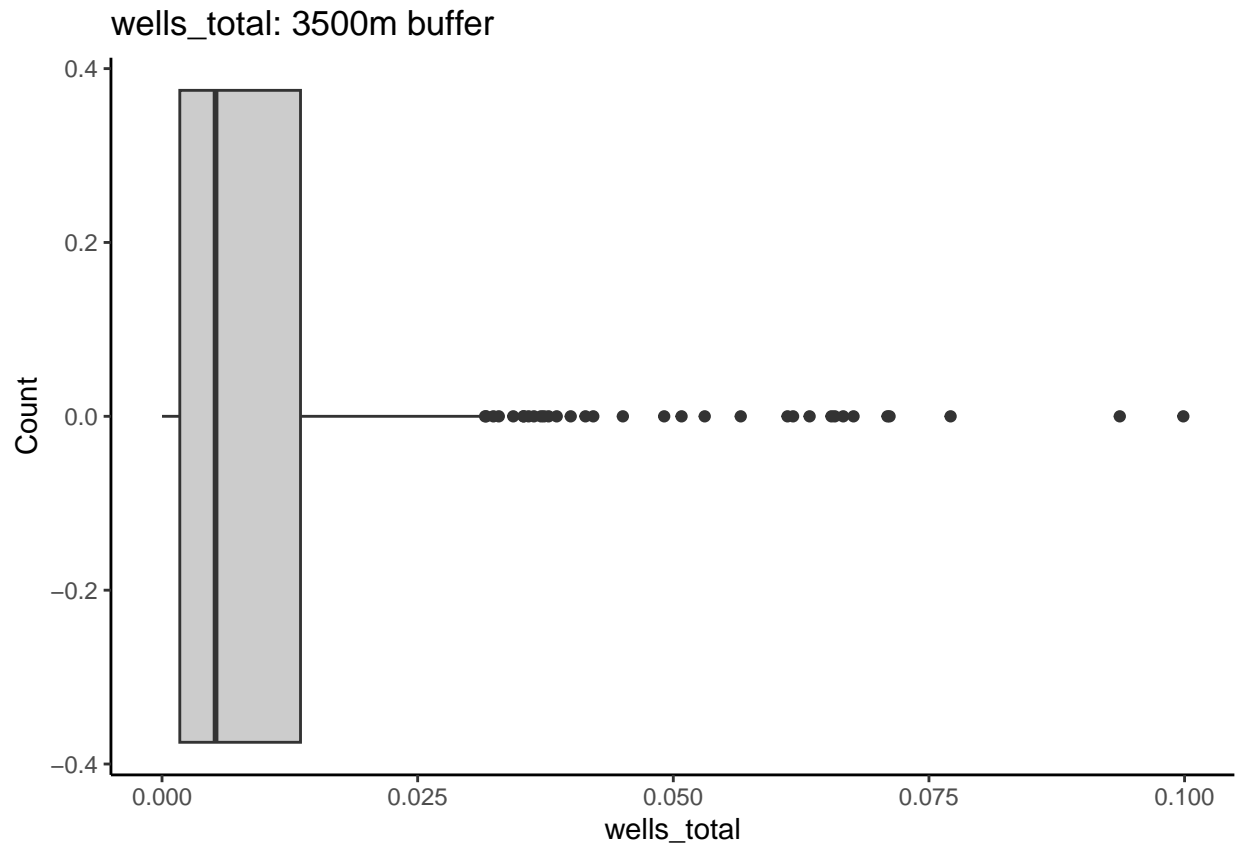




```
##  
## $wells_inactive
```



```
##  
## $wells_total
```



### 3. Correlation

Now let's look at correlation among variables. We have a ton of variables to look at so for now we'll export to Excel.

```
# Histograms for each buffer size
corr_data <- purrr::map(unique(covs_to_explore$buffer_dist),
  ~ covs_to_explore %>%

  filter(buffer_dist == .x) %>%

  # Remove the response/site columns
  select(9:ncol()) %>%

  # pairwise complete observations right now since
  # I'm not sure how to manage all the NAs in the configuration metrics yet
  corrr::correlate(., use = "pairwise.complete.obs",
    method = "spearman",
    diagonal = NA,
    quiet = TRUE) %>%

  as_tibble() %>%

  # Remove the bottom triangle to keep things tidy
  mutate(across(everything(), ~ if_else(row_number() >=
    ↪ min(which(is.na(.))), NA, .))) %>%
```

```

      # Get the triangle shaped properly
      select(rev(names(.))) %>%

      # Term at the beginning
      relocate(term)

    ) %>%

    # Set the names of each list element to the buffer they represent
    set_names(paste0(unique(covs_to_explore$buffer_dist), "m"))

    # Nifty function here will export each list element as a separate tab in Excel :)
    writexl::write_xlsx(corr_data, "./data/processed/OSM_all_covariates_correlation.xlsx")

```

Let's also just look at the configuration metrics in a more compact format

```

# Histograms for each buffer size
corr_data_config <- purrr::map(unique(covs_to_explore$buffer_dist),
  ~ covs_to_explore %>%

    filter(buffer_dist == .x) %>%

    select(contains("landscape") | contains("nonanthro")) %>%

    # pairwise complete observations right now since I'm not sure how to
    ↪ manage all the NAs in the configuration metrics yet
    corrr::correlate(., use = "pairwise.complete.obs",
      method = "spearman",
      diagonal = NA,
      quiet = TRUE) %>%

    # Same tidy and reformatting as above
    as_tibble() %>%

    mutate(across(everything(), ~ if_else(row_number() >=
      ↪ min(which(is.na(.))), NA, .))) %>%

    select(rev(names(.))) %>%

    relocate(term)

  ) %>%

  # set names to buffer distance
  set_names(paste0(unique(covs_to_explore$buffer_dist), "m"))

# Export to excel
writexl::write_xlsx(corr_data_config,
  ↪ "./data/processed/OSM_all_config_covariates_correlation.xlsx")

```

## 4. Summary table of covariates

This will need to be tweaked at the end of the process to include all variables **at the appropriate spatial scales corresponding to the top results from the dredge**.

```

# This is the formatted table for the paper with max/min values, etc.
covariate_table <-

bind_cols(

  # natural data scaled by natural buffer
  covs_to_explore %>%
    filter(buffer_dist == 100) %>%
    select(fire_0_15, contains("lc_")),

  # composition data scaled by composition buffer
  covs_to_explore %>%
    filter(buffer_dist == 4250) %>%
    select(harvest_0_15, harvest_total, osm_industrial, pipe_trans, roads, seismic, wells_inactive,
      ↪ wells_active),

  # configuration data
  covs_to_explore %>%
    filter(buffer_dist == 2250) %>%
    select(nonanthro_ed, landscape_mesh, landscape_shei, nonanthro_cohesion, cfi_site)
) %>%

# Replace NAs with zeros.
# This should be redundant code at this point and could probably be deleted
mutate(across(everything(), ~ replace_na(.x, 0))) %>%

# z used as a unique delimiter `z` to separate during pivot wider.... hacky but it works
summarise(
  across(everything(), list(
    zmean = ~mean(.),
    zmin = ~min(.),
    zmax = ~max(.),
    zmedian = ~median(.),
    zq20 = ~quantile(., probs = 0.05),
    zq80 = ~quantile(., probs = 0.95)
  ))
) %>%

pivot_longer(cols = everything(), names_to = c("Covariate", ".value"), names_sep = "_z") %>%

# Paste into one string (as in the table) so we don't have to do it in Excel
mutate(`Median (5%, 95% quantiles)` = paste0(round(median, 2),
  " (", round(q20, 2),
  "-",
  round(q80, 2), ")"),

  # ...and again
  `Mean (min, max)` = paste0(round(mean, 2),
    " (", round(min, 2),
    "-",
    round(max, 2), ")"),

) %>%

# Remove trails if it exists
filter(Covariate != 'trails') %>%

# REmove unformatted columns
select(-mean, -min, -max, -median, -q20, -q80) %>%

# Blank description column we can export later

```

```

mutate(`Modelled Scale` = "",
      Description = "") %>%

# Replace the covariate names with these pretty ones that we stored in a table
left_join(read_csv("./tables/OSM_all_covariates_formatted_names.csv"),
          by = "Covariate") %>%

# Swap out the names
mutate(UglyName = Covariate,
      Covariate = PrettyName, .keep = "unused")

# End the table
writexl::write_xlsx(covariate_table, "./tables/OSM_all_covariates_summary.xlsx")

```

End script.