Overview of Data Sources for Coastal Water Quality

Francois van Schalkwyk

2025-04-03

Make sure that you are working in the outfalls project, found in the working directory.

# 1 Sample Frequency

## 1.1 Coastal Management Branch Excel Spreadsheets

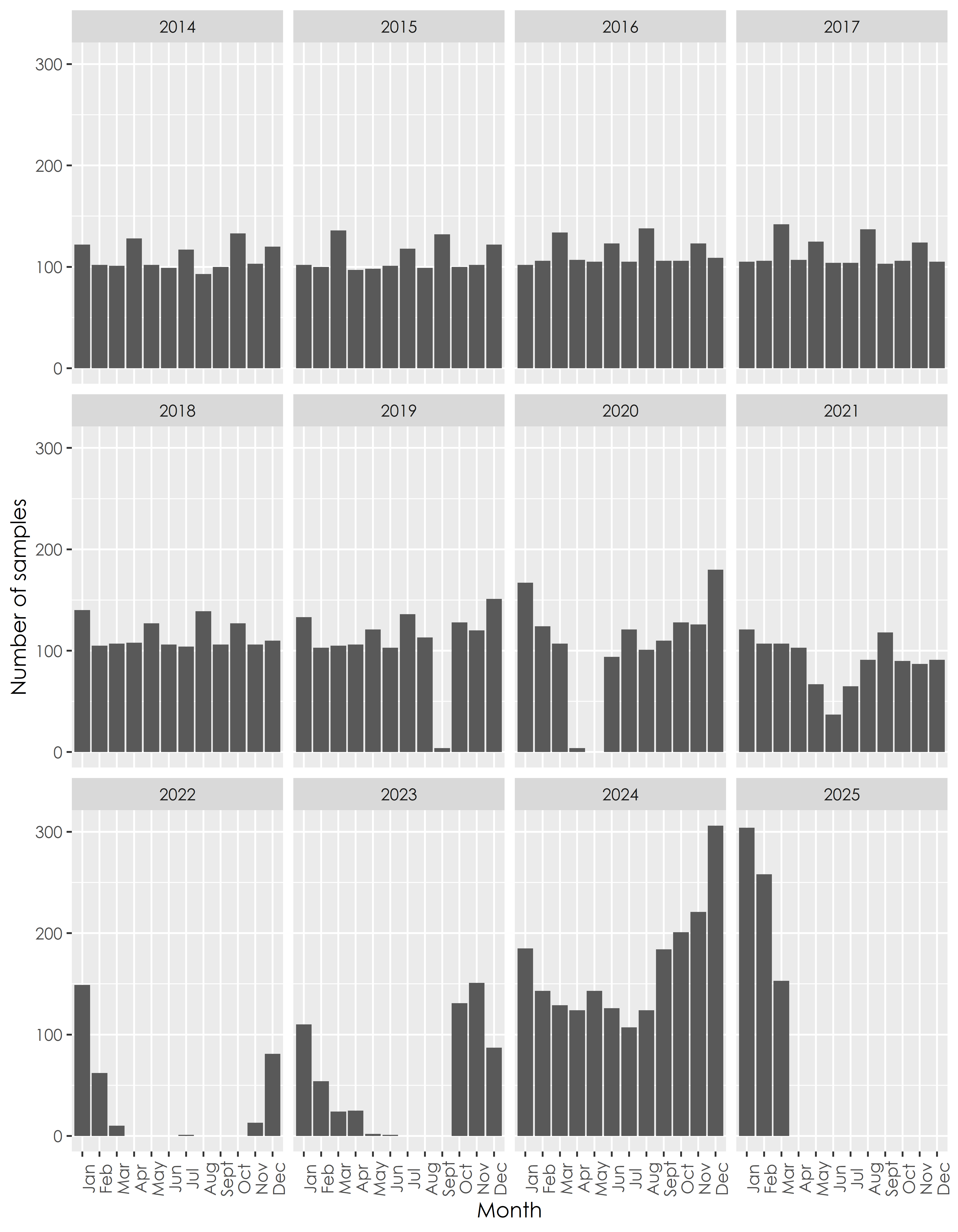


Figure 1.1: Number of samples received per month from data received from Coastal Management Branch

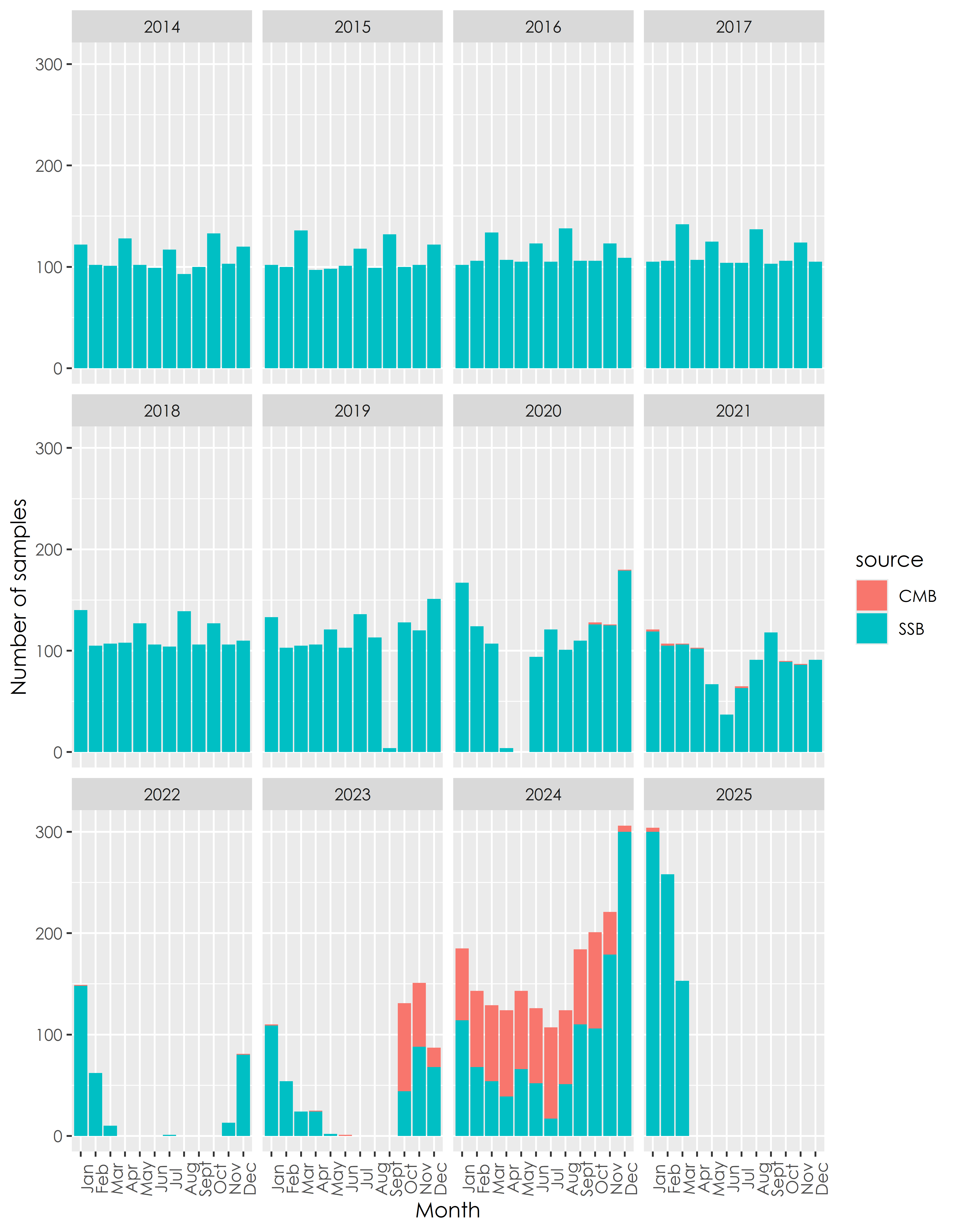


Figure 1.2: Number of samples received per month from data received from Coastal Management Branch

## 1.2 Data received from LIMS dataset

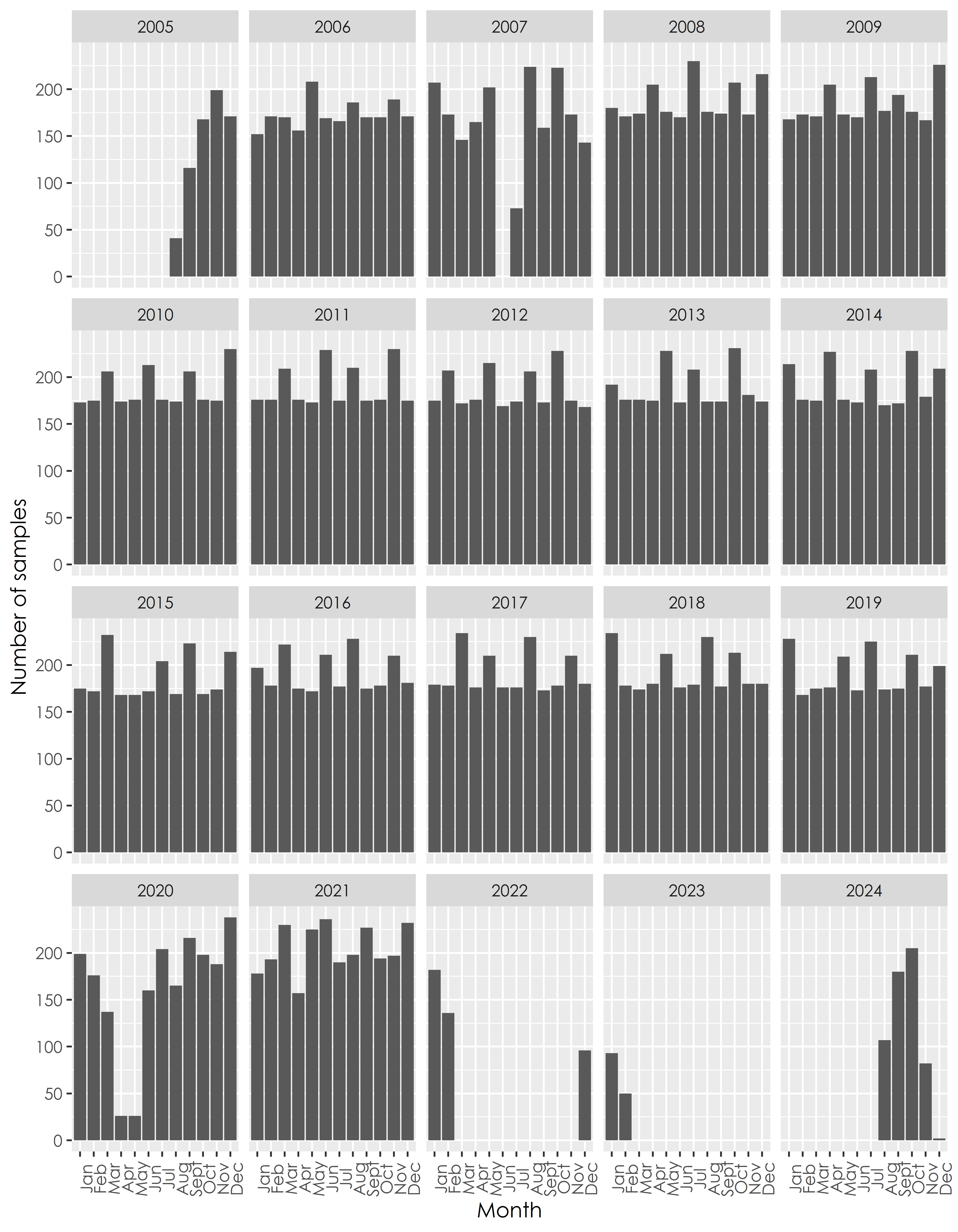


Figure 1.3: Number of samples received per month from 2005 to 2024 according to the LIMS dataset

## 1.3 Overlap between CMB and LIMS datasets

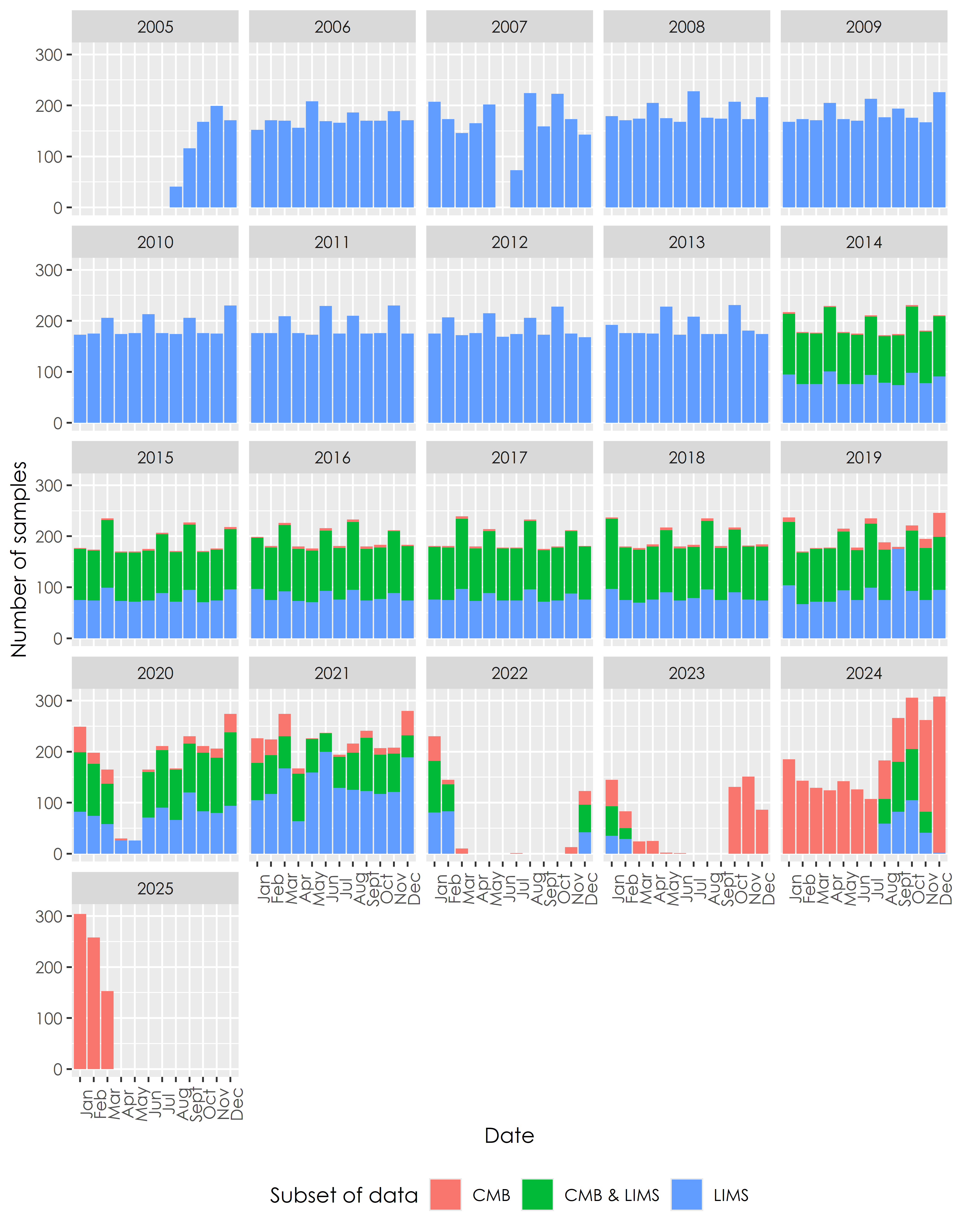


Figure 1.4: Overlap of data between CMB and LIMS datasets

Key points to note: - There are data in the LIMS dataset that are not in the CMB spreadsheets throughout the range of the data - There is also a small amount of data in the Excel spreadsheets that are not in the LIMS data, and this increases from 2023 onwards as the SABS data were populated into the Excel spreadsheets

## 1.4 Overlap between LIMS data and SABS data

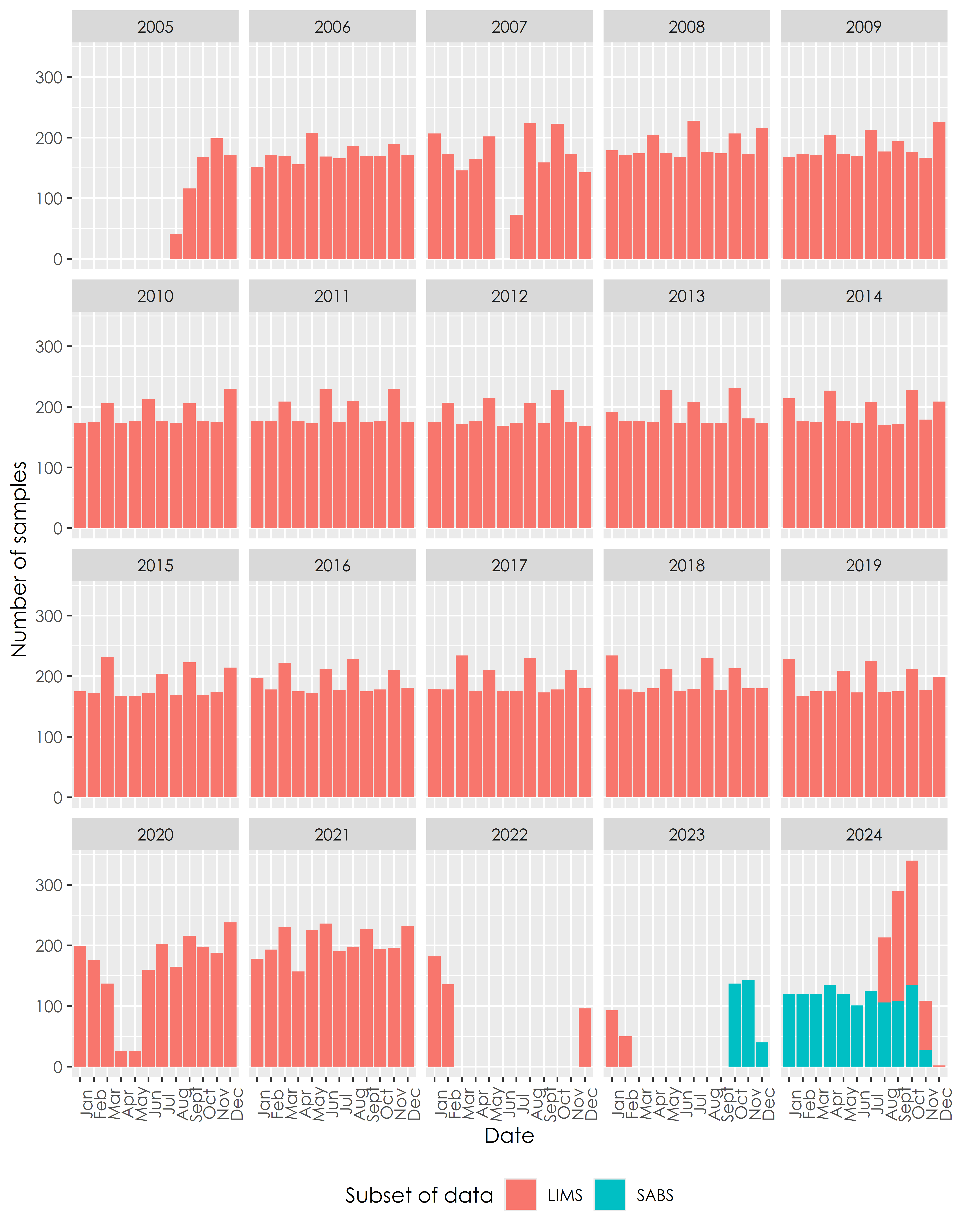


Figure 1.5: Overlap of data between CMB and LIMS datasets

Key points to note: - There is no overlap in data between the SABS data from 2024 and 2025, and the LIMS data - The LIMS data in 2024 and 2025 are therefore completely made up of SSB data, and the results are typical of Enterolert data - This makes it highly likely that these data come from Enterolert results from within the SSB

## 1.5 All data sources shown together

### 1.5.1 All data sources from 2005 to 2025

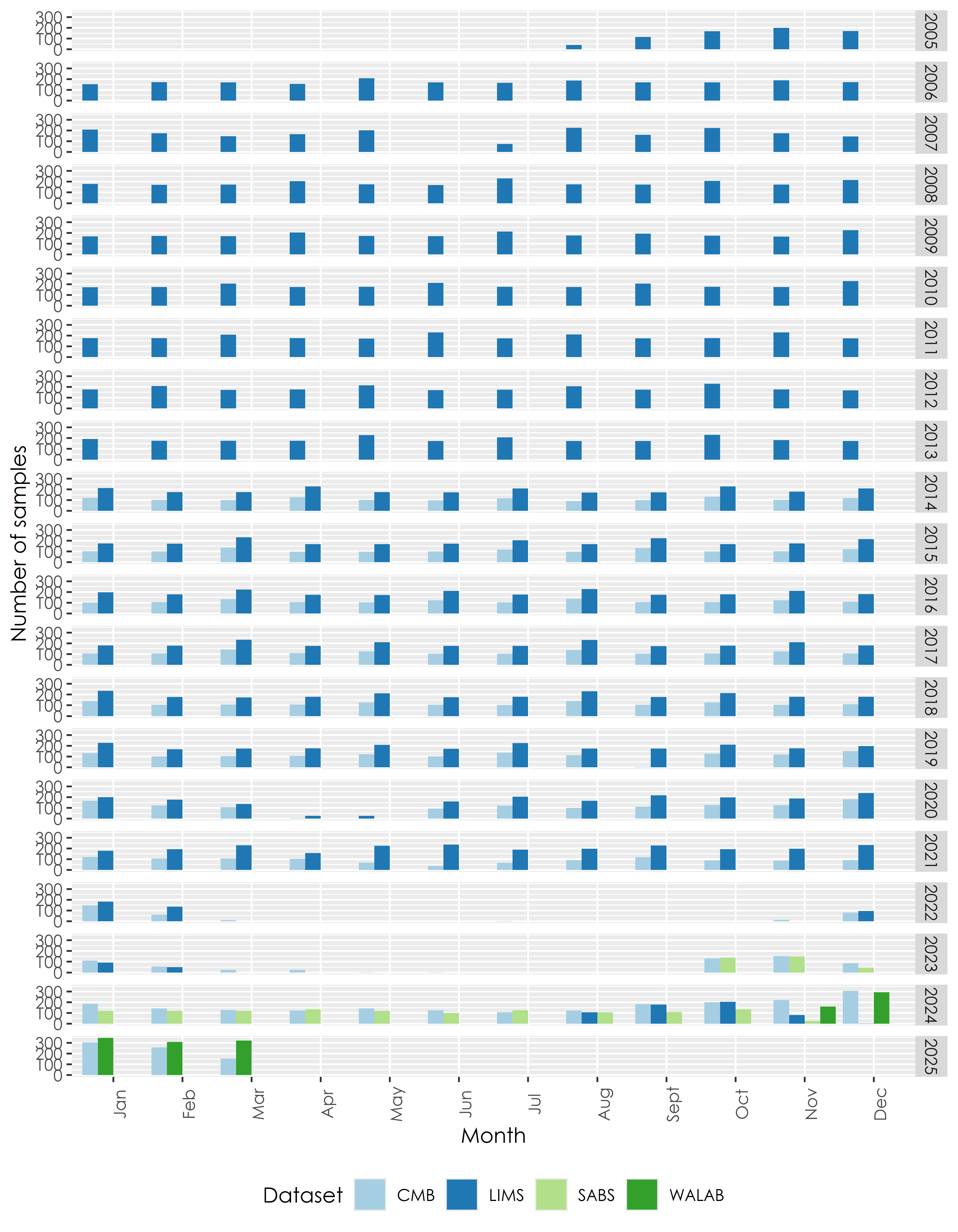


Figure 1.6: Number of samples per month for all datasets

### 1.5.2 All data sources from 2020 to 2025

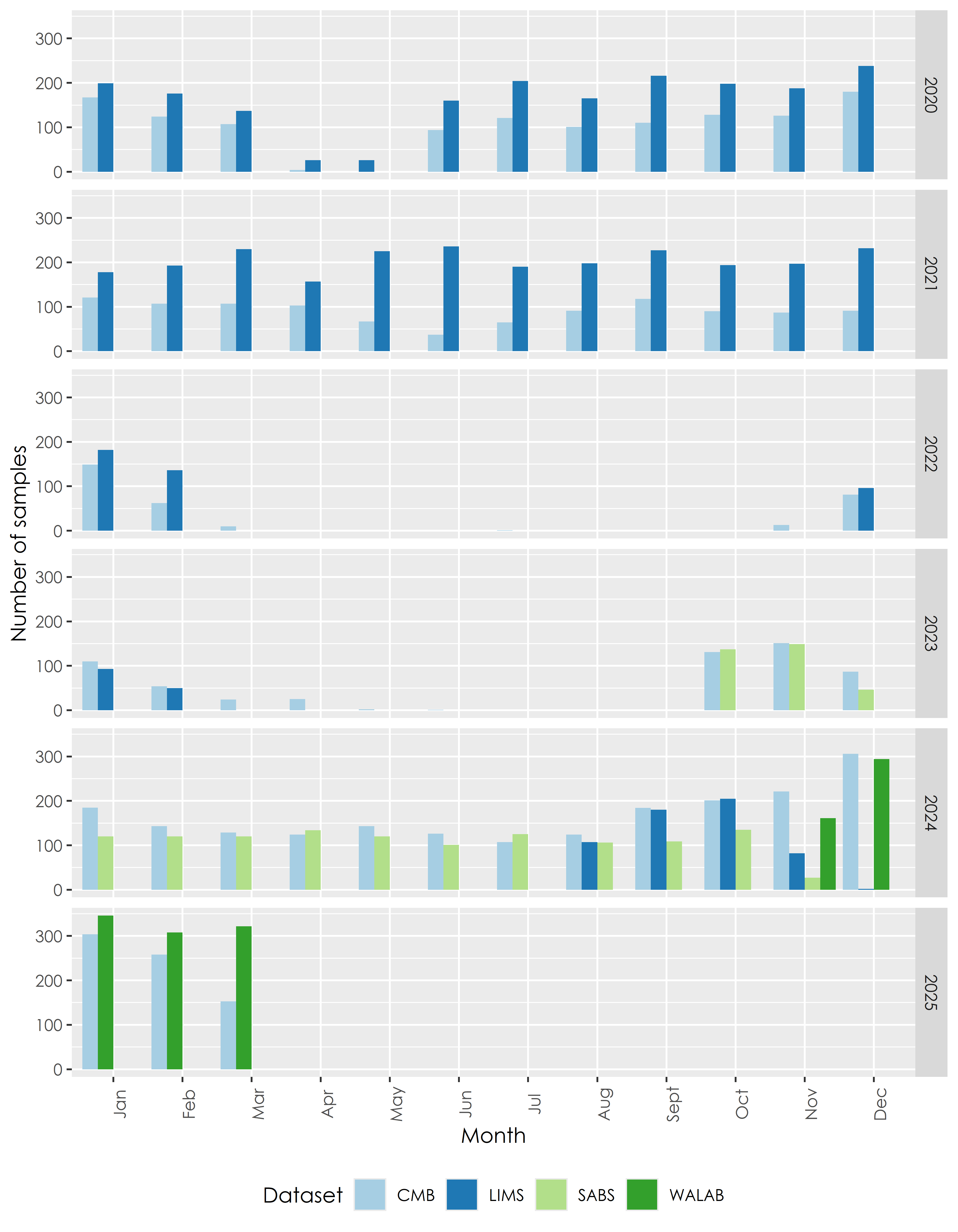


Figure 1.7: Number of samples per month for all datasets, for the last five years

Key points to note: - Note the sparsity of the data in 2022 and 2023, with the AL Abbot data situated in the middle of this period - There are differences between CMB spreadsheets and SABS data in 2023 and 2025 - There are differences between CMB spreadsheets and WALAB data in 2024 and 2025

# 2 Results

## 2.1 Results for all datasets

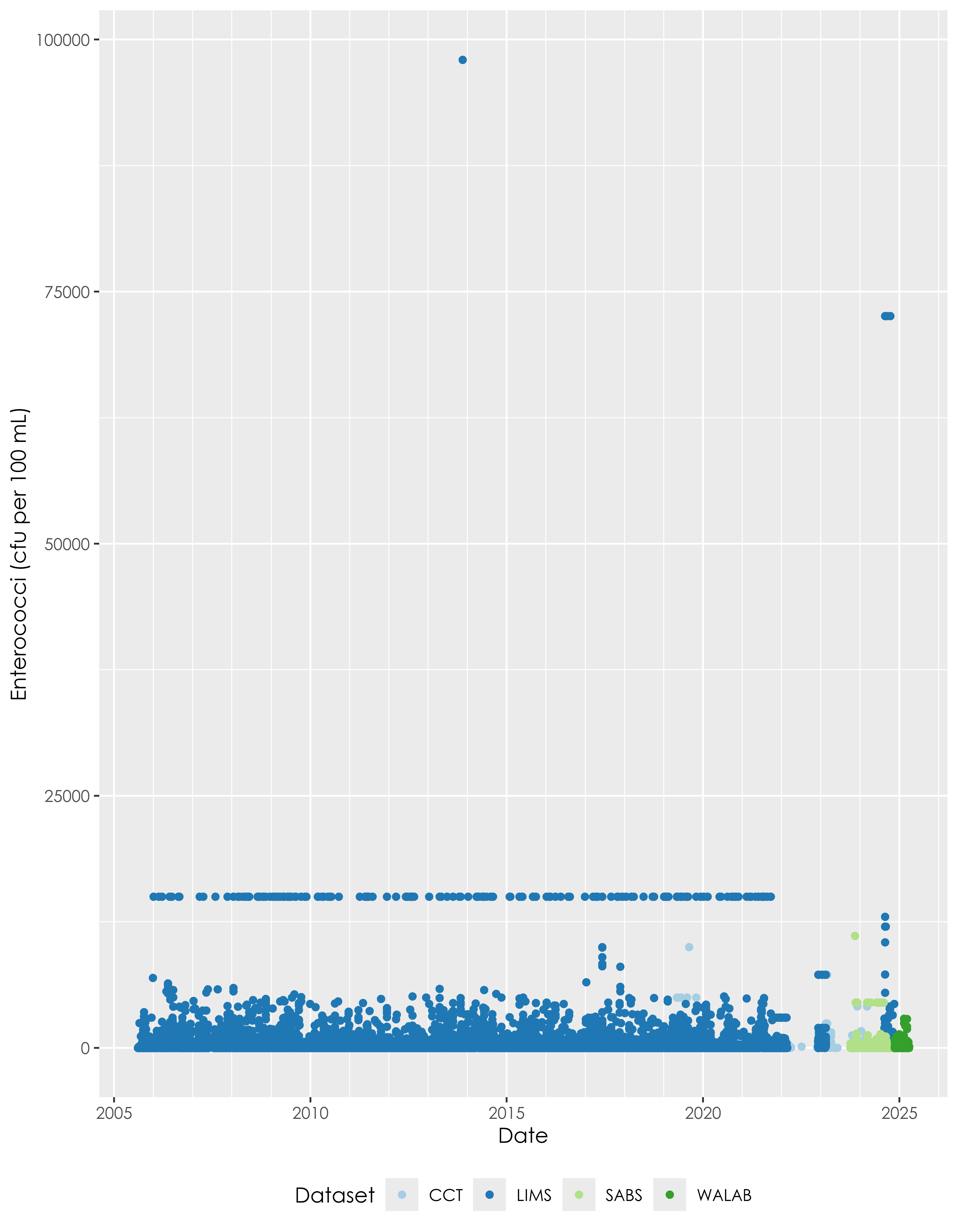


Figure 2.1: Results for all datasets

## 2.2 Outlier data in August 2024 from LIMS data using Enterolert

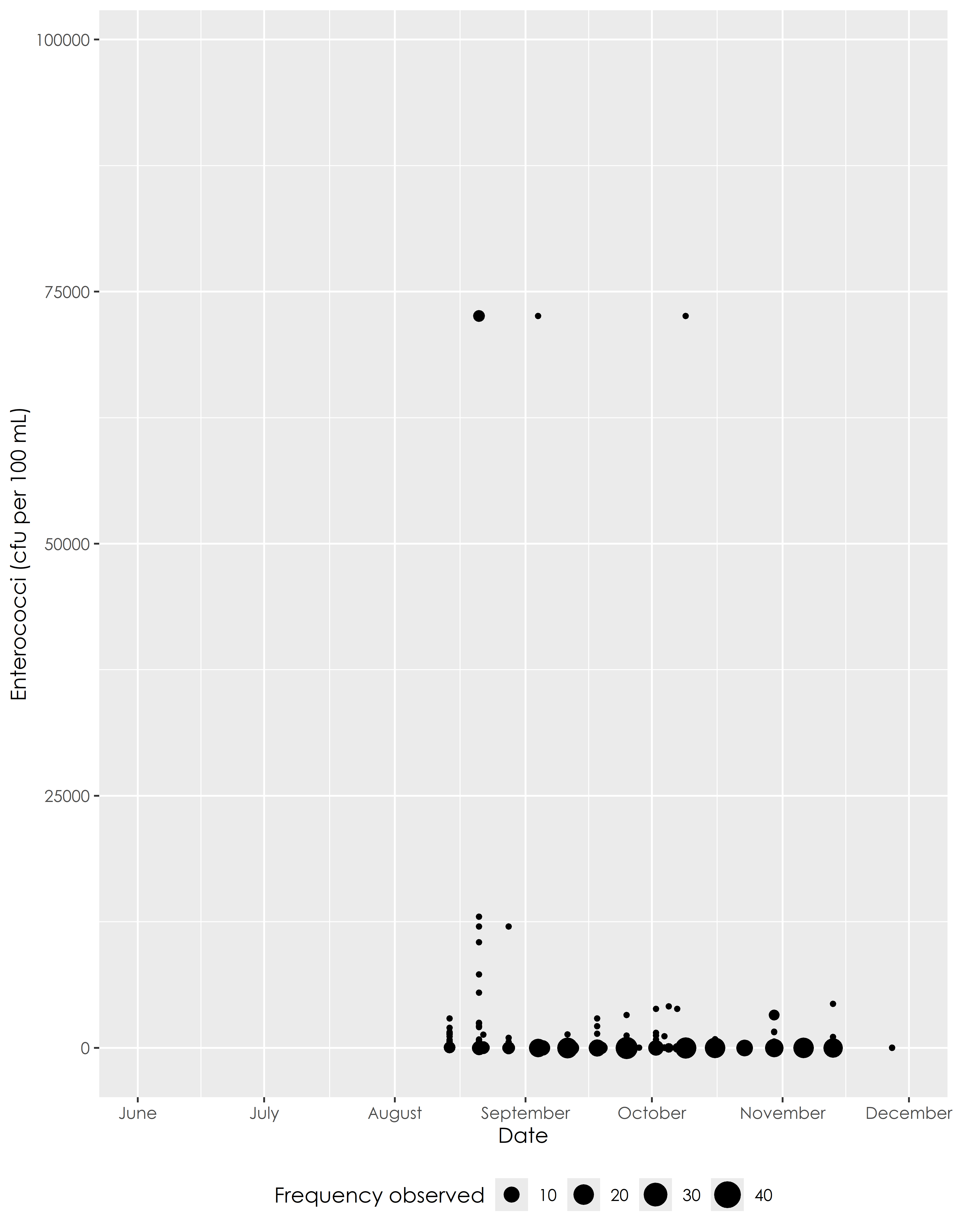


Figure 2.2: Outlier data in August 2024 from the LIMS data set

Table 2.1: Very high values according to LIMS dataset on 21 August 2024

| Sample Date | Site ID | Censored Value (cfu per 100 mL) | Numeric Value (cfu per 100 mL) |
| --- | --- | --- | --- |
| 2024-08-21 | CS04 | >24196 | 72588 |
| 2024-08-21 | CS08 | >24196 | 72588 |
| 2024-08-21 | CS09 | >24196 | 72588 |
| 2024-08-21 | CS12 | >24196 | 72588 |

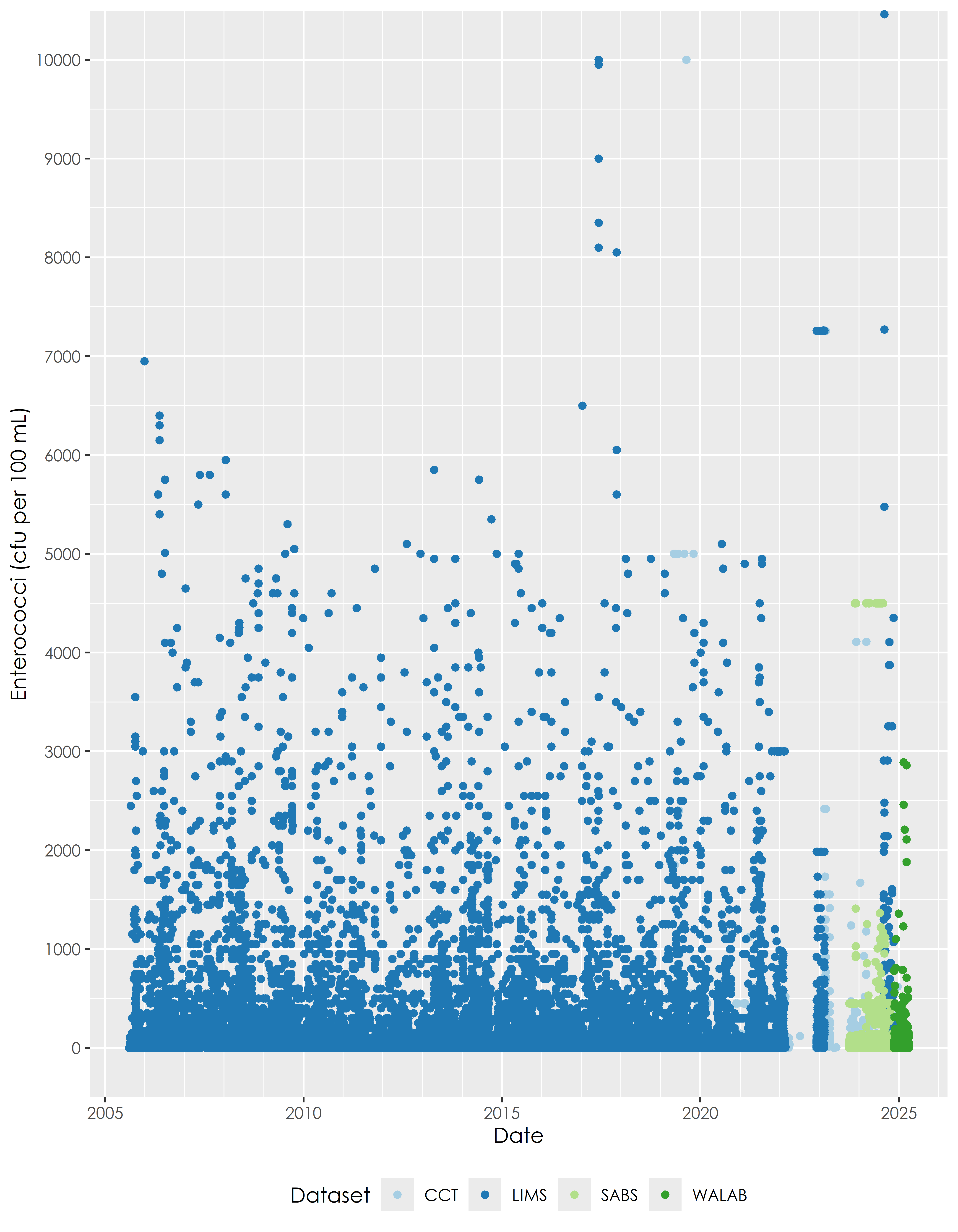


Figure 2.3: Results for all datasets excluding outlier values

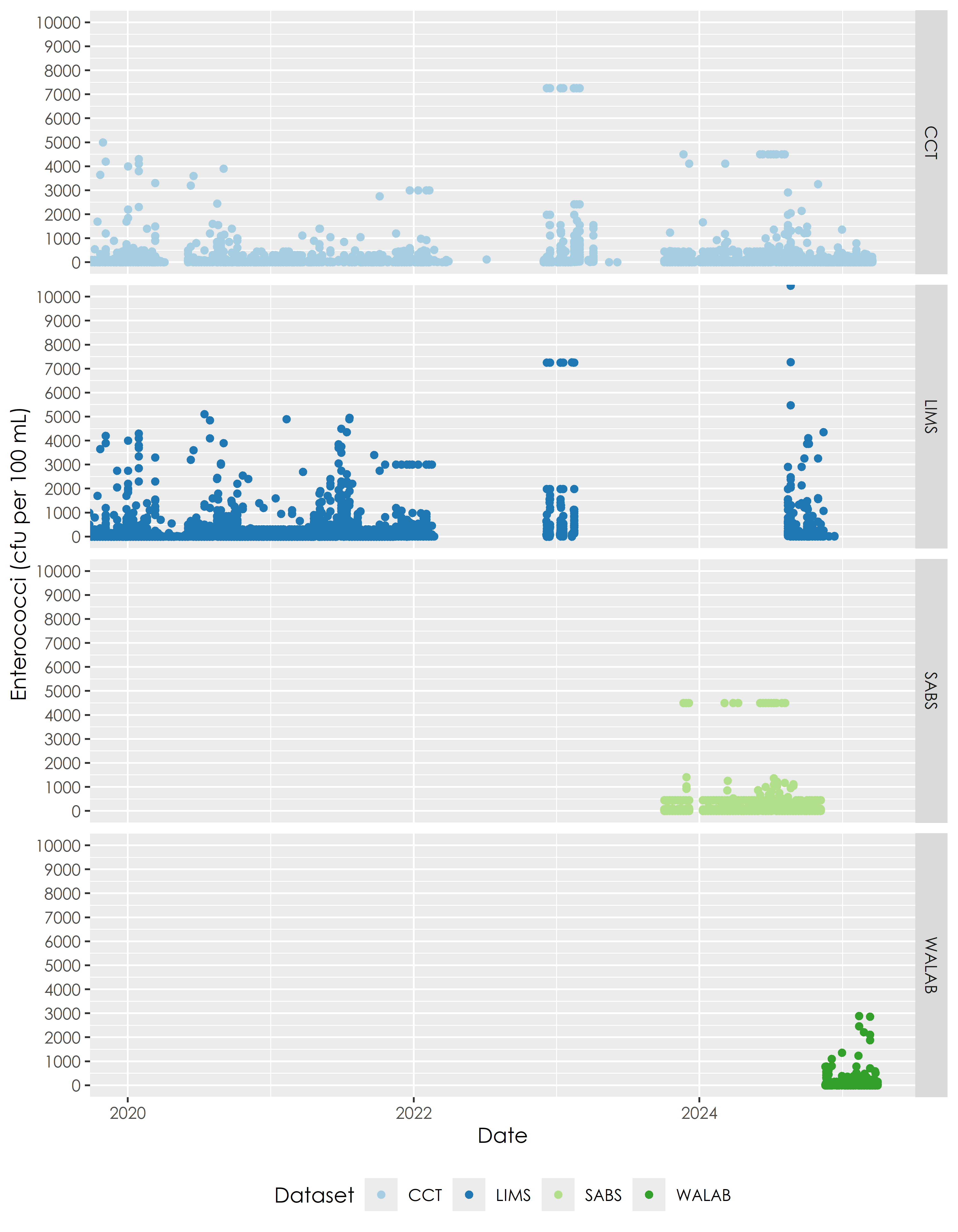


Figure 2.4: Results for all datasets from 2020 to 2025

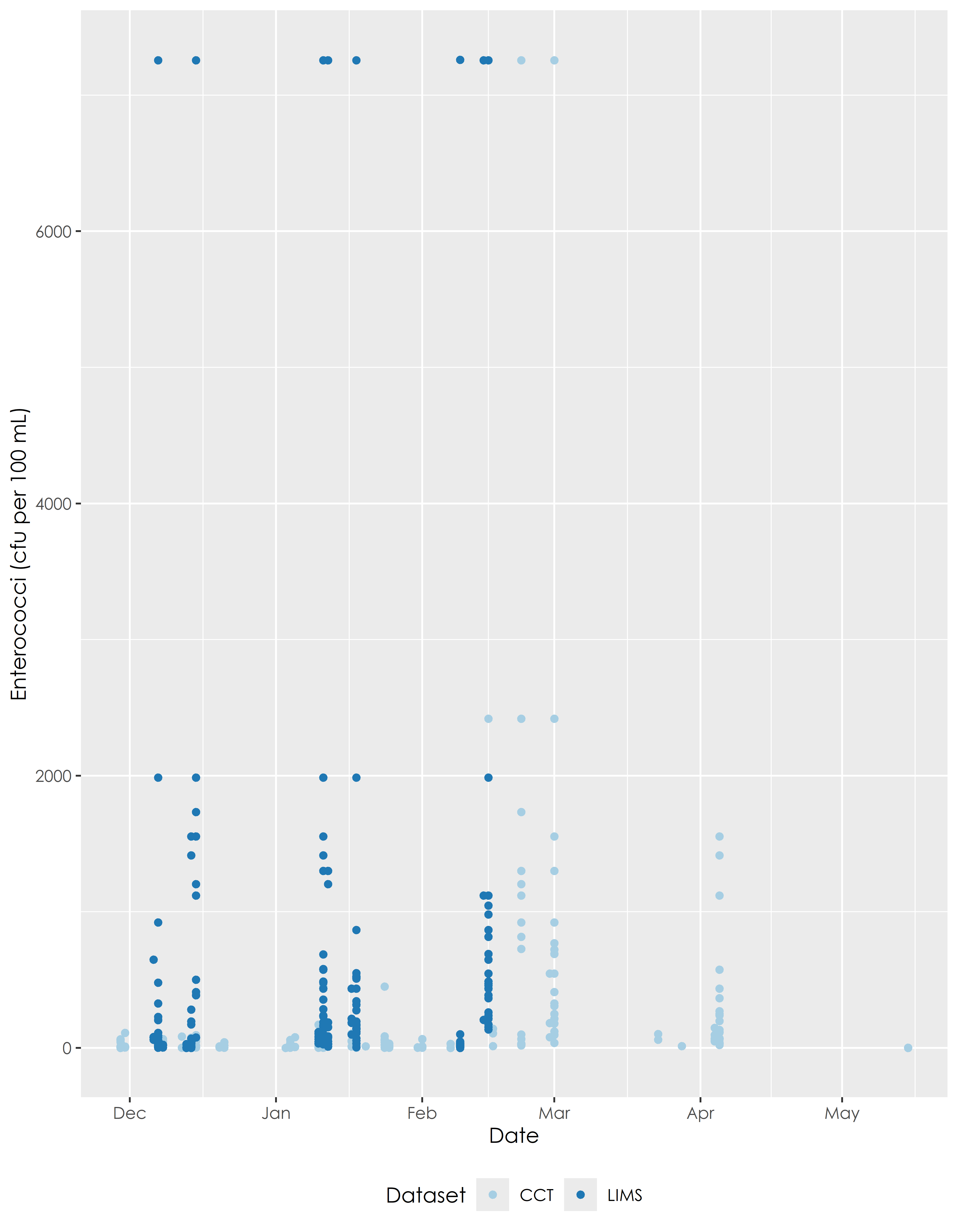


Figure 2.5: Results for AL Abbot data between 2022 and 2023

# 3 Sites

# CCT and SABS  
dplyr::setdiff(cct\_sites, sabs\_sites) # 52 sites

## # A tibble: 52 × 1  
## site\_id  
## <chr>   
## 1 CN06C   
## 2 CN10   
## 3 CN11   
## 4 CN14   
## 5 CN16I   
## 6 CN18I   
## 7 CN19I   
## 8 CN20I   
## 9 CN21   
## 10 CN30   
## # ℹ 42 more rows

dplyr::setdiff(sabs\_sites, cct\_sites) # 11 sites

## # A tibble: 11 × 1  
## site\_id  
## <chr>   
## 1 CN04   
## 2 CN38   
## 3 CN42   
## 4 CS12   
## 5 HB13   
## 6 ICS01   
## 7 ICS03   
## 8 ICS06   
## 9 ICS08   
## 10 ICS09   
## 11 ICS14

# CCT and WALAB  
dplyr::setdiff(cct\_sites, walab\_sites) # 2 sites

## # A tibble: 2 × 1  
## site\_id  
## <chr>   
## 1 HB09   
## 2 XCS16

dplyr::setdiff(walab\_sites, cct\_sites) # 33 sites

## # A tibble: 33 × 1  
## site\_id  
## <chr>   
## 1 CN04   
## 2 CN05   
## 3 CN08O   
## 4 CN12O   
## 5 CN16A   
## 6 CN16O   
## 7 CN18O   
## 8 CN19O   
## 9 CN29   
## 10 CN38   
## # ℹ 23 more rows

# SABS and LIMS  
dplyr::setdiff(sabs\_sites, lims\_sites) # 11 sites

## # A tibble: 11 × 1  
## site\_id  
## <chr>   
## 1 CN38   
## 2 CN41   
## 3 CN42   
## 4 CS37   
## 5 HB13   
## 6 ICS01   
## 7 ICS03   
## 8 ICS06   
## 9 ICS08   
## 10 ICS09   
## 11 ICS14

dplyr::setdiff(lims\_sites, sabs\_sites) # 79 sites

## # A tibble: 79 × 1  
## site\_id  
## <chr>   
## 1 CN02A   
## 2 CN03   
## 3 CN05   
## 4 CN05A   
## 5 CN06A   
## 6 CN06C   
## 7 CN08I   
## 8 CN08O   
## 9 CN10   
## 10 CN11   
## # ℹ 69 more rows

# LIMS and WALAB  
dplyr::setdiff(lims\_sites, walab\_sites) # 19 sites

## # A tibble: 19 × 1  
## site\_id  
## <chr>   
## 1 CN02A   
## 2 CN03   
## 3 CN05A   
## 4 CN06A   
## 5 CN08I   
## 6 CN12B   
## 7 CN15   
## 8 CN20O   
## 9 CN32   
## 10 CS01   
## 11 CS05   
## 12 CS08   
## 13 CS12   
## 14 CS24E   
## 15 CS24W   
## 16 XCN03A   
## 17 XCN15   
## 18 XCS16   
## 19 XCS33

dplyr::setdiff(walab\_sites, lims\_sites) # 23 sites

## # A tibble: 23 × 1  
## site\_id  
## <chr>   
## 1 CN34   
## 2 CN35   
## 3 CN36   
## 4 CN37   
## 5 CN38   
## 6 CN39   
## 7 CN40   
## 8 CN41   
## 9 CN42   
## 10 CS30   
## # ℹ 13 more rows

# CCT and LIMS  
dplyr::setdiff(lims\_sites, cct\_sites) # 44 sites

## # A tibble: 44 × 1  
## site\_id  
## <chr>   
## 1 CN02A   
## 2 CN03   
## 3 CN04   
## 4 CN05   
## 5 CN05A   
## 6 CN06A   
## 7 CN08I   
## 8 CN08O   
## 9 CN12B   
## 10 CN12O   
## # ℹ 34 more rows

dplyr::setdiff(cct\_sites, lims\_sites) # 17 sites

## # A tibble: 17 × 1  
## site\_id  
## <chr>   
## 1 CN34   
## 2 CN35   
## 3 CN36   
## 4 CN37   
## 5 CN39   
## 6 CN40   
## 7 CN41   
## 8 CS36   
## 9 CS37   
## 10 CS38   
## 11 CS41   
## 12 CS42   
## 13 CS44   
## 14 CS45   
## 15 CS46   
## 16 HB09   
## 17 XCS34

# SABS and WALAB  
dplyr::setdiff(sabs\_sites, walab\_sites) # 7 sites

## # A tibble: 7 × 1  
## site\_id  
## <chr>   
## 1 CS12   
## 2 ICS01   
## 3 ICS03   
## 4 ICS06   
## 5 ICS08   
## 6 ICS09   
## 7 ICS14

dplyr::setdiff(walab\_sites, sabs\_sites) # 79 sites

## # A tibble: 79 × 1  
## site\_id  
## <chr>   
## 1 CN05   
## 2 CN06C   
## 3 CN08O   
## 4 CN10   
## 5 CN11   
## 6 CN12O   
## 7 CN14   
## 8 CN16A   
## 9 CN16I   
## 10 CN16O   
## # ℹ 69 more rows

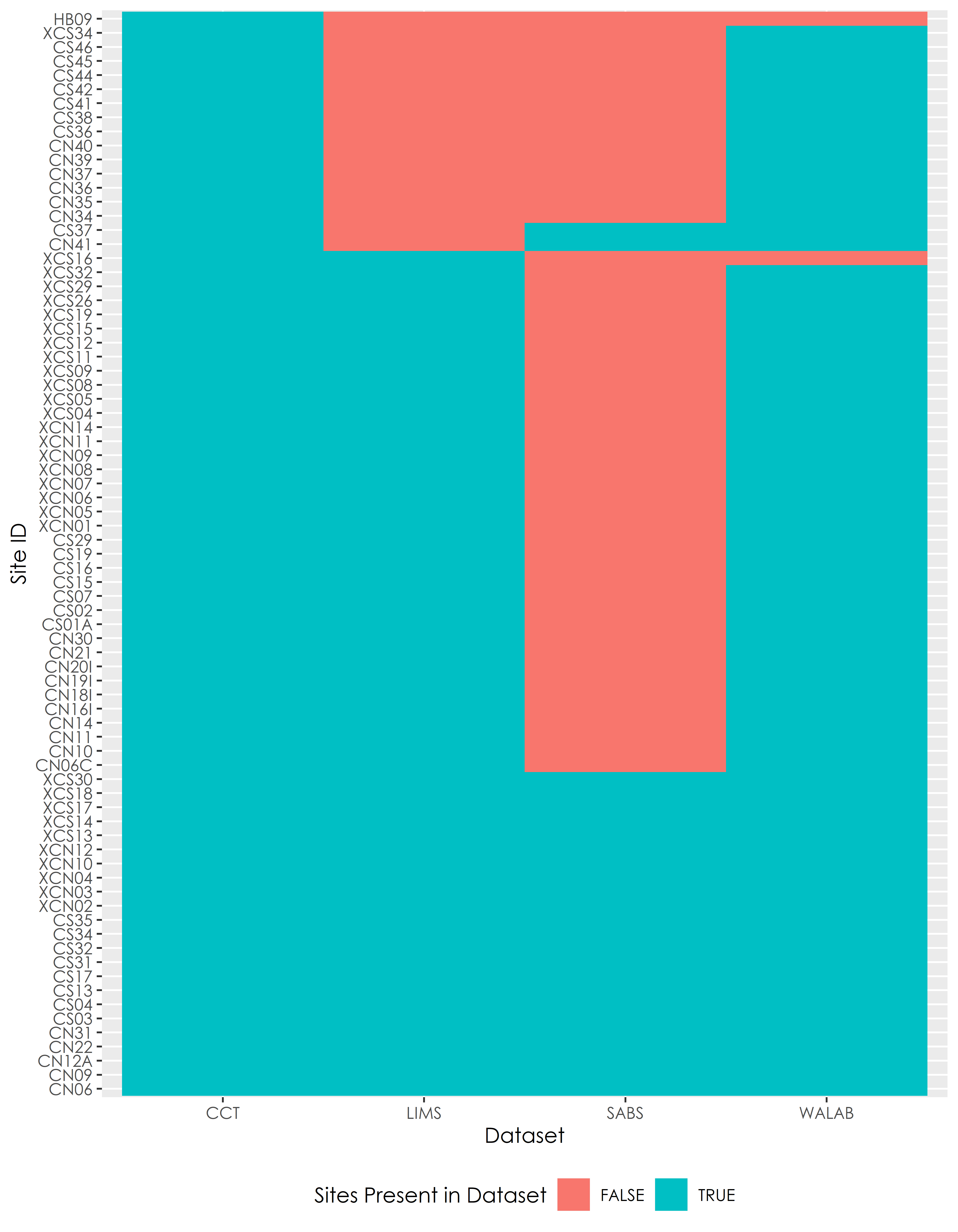


Figure 3.1: Sites that are present in CCT dataset

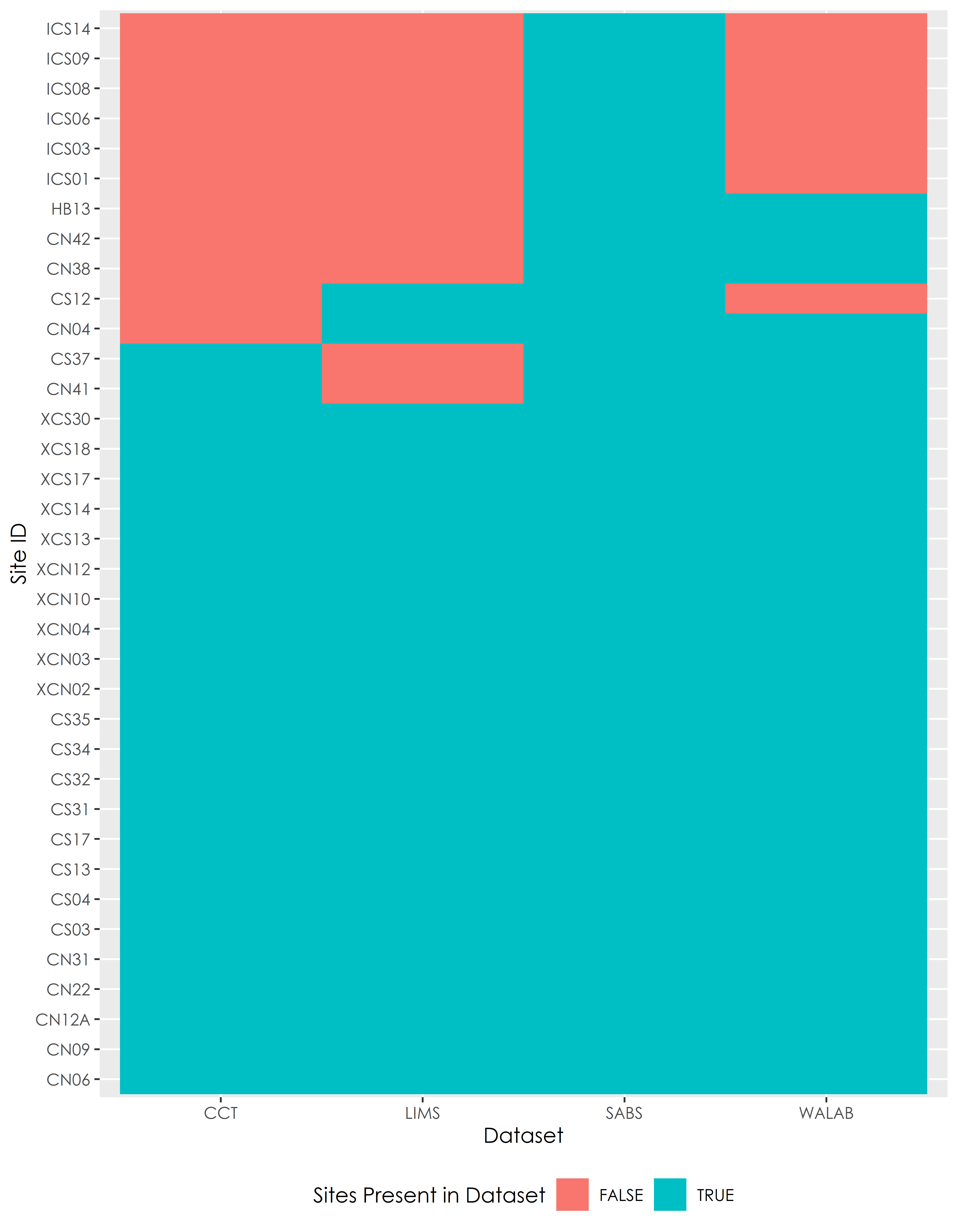


Figure 3.2: Sites that are present in SABS dataset

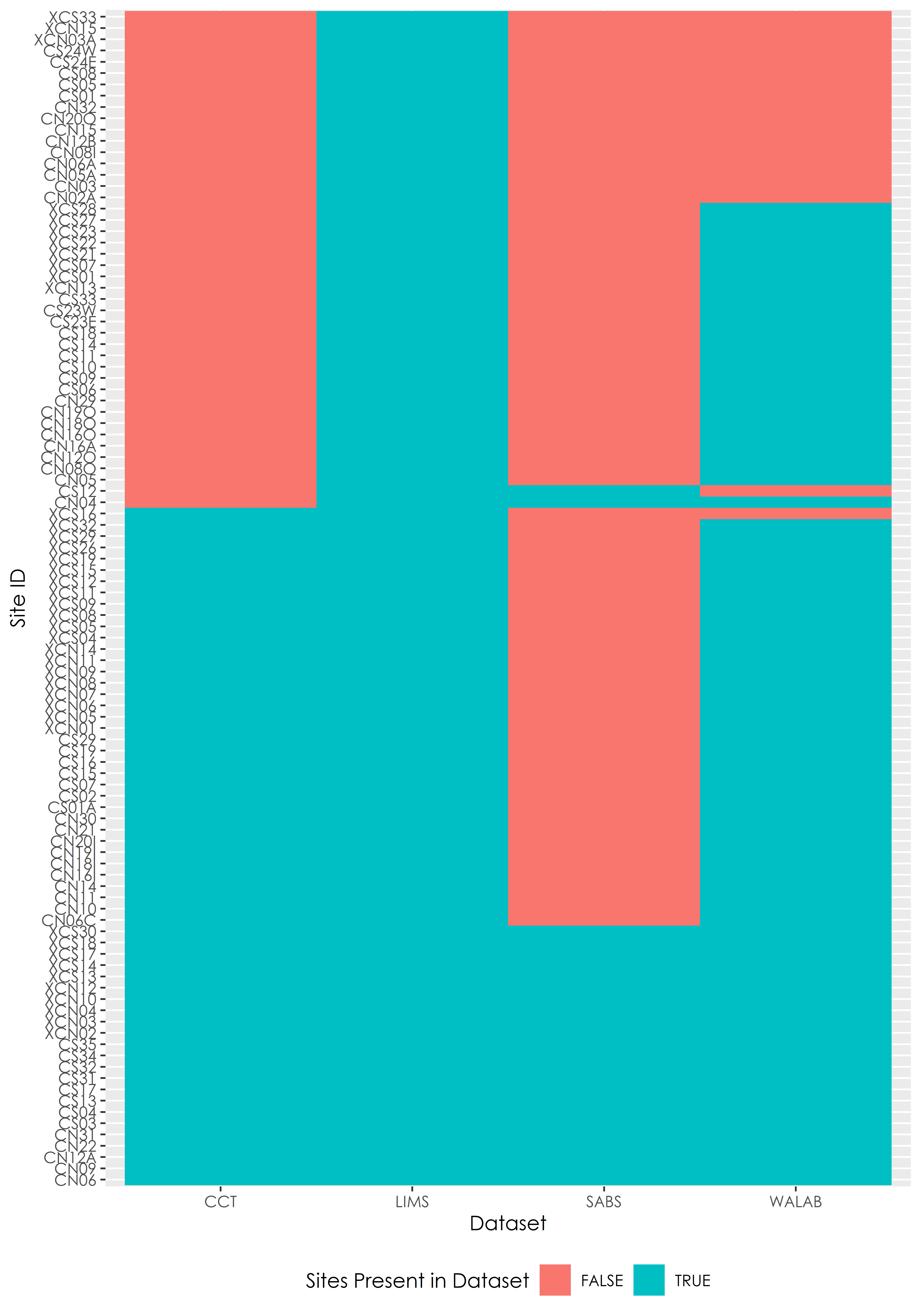


Figure 3.3: Sites that are present in LIMS dataset

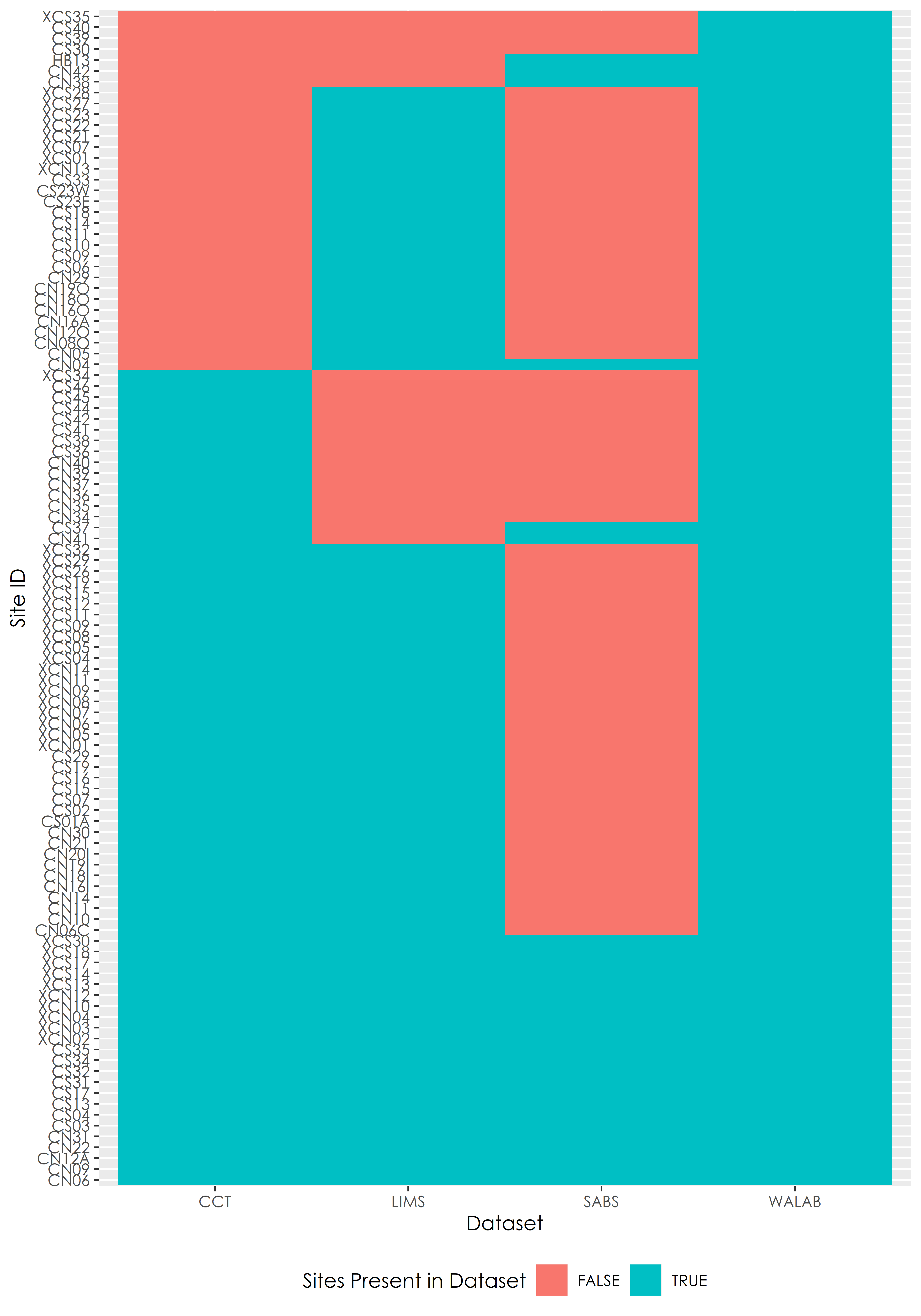


Figure 3.4: Sites that are present in WALAB dataset