PFAS in North Carolina

https://github.com/kn134/EDAFinal_PFAS.git

Karly Nocera and Tay Holliday

4/26/2021

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

The purpose of this study was to synthesize and analyze the findings of local water quality samples measuring a group of emerging contaminants, PFAS (per- and polyfluoroalkyl substances). The two datasets from DEQ (2018 Emerging Compounds Monitoring Reports of various watersheds and public water supply (PWS) reservoirs and the 2019 wastewater treatment plant (WWTP) samples) were chosen as the only publicly-available datasets on local PFAS levels. The sample locations and variables were pre-determined by the study, but these datasets reflect locations and variables pertaining to the levels of various PFAS analytes in the drinking water and water bodies associated with 2.7 million North Carolina residents.

Our study aims to explore and analyze this data in order to begin building the larger story of local contamination and identify areas of concern that need further investigation. To do so, we answered the following questions:

1. Occurence

- What are the most abundant analytes present?
- Which sites have analytes exceeding health advisory levels?
- What chain lengths are most abundant and present?
- What portion of total PFAS are PFOA and PFOS?

2. Distribution

- Are there bodies of water with concerningly high levels of PFAS that aren't commonly discussed?
- How does total PFAS change over time?

3. Unreported / Further Investigation

- What is the highest reporting limit for each analyte?
- How many samples had reporting limits above health advisory levels?
- What percentage of total unreported results were below health advisory limits?

2 Dataset Information

The two datasets were downloaded from North Carolina DEQ: 2018 Analytical Results for PFAS Screening of Select Public Water Supply (PWS) Reservoirs and 2019 publicaly owned utilities with pretreatment programs (POTWs) and industrial dischargers with state permits. Respective to their type of site, both datasets contained site location, sample date, PFAS analyte levels (nanograms per liter (ng/L) or equivalent parts per trillion (ppt)), and lab qualifiers/reporting limits. To substantiate this analysis, we incorporated a manually created csv file identifying the latitude and longitude of sample sites, and researched analytes' chain length (short or long).

Due to the nature of their original format, both datasets needed extensive cleaning prior to importing to R and further exploration and analysis. The POTW data was originally in a PDF and first needed to be converted (using SmallPDF tool) to an excel spreadsheet. These values were then manually moved intoa dataset layout (column headings, rows for each analyte sampled). The PWS dataset was already a csv dataset. For both, analytes with dashes (-) needed to be replaced with a different character (.) in order to read into R and column names were changed to be consistent with other. The lab qualifiers were split so that there was a unique column of numeric reporting limits and in PWS the "results" value was deleted because it was actually the reporting limit and not a result, despite being originally in that column (indicated with corresponding lab qualifier characters).

Once these three datasets were imported into R Studio, both POTW and PWS needed the "sample date" column formatted as a date class and unnecessary columns were removed. To pivot each wider for future use, their reporting limit column was removed and then each pivoted wider so that each analyte became a column with its corresponding parts per trillion measurement. We then mutated both datasets to create columns for Total PFAS, the sum of PFOA and PFOS (due to health advisory regulation), the sum of all short-chained PFAS, and the sum of all long-chained PFAS.

To create a column for chain length in the original long datasets, we used a function to determine if each analyte was short or long, adding the result in the column. Then, both were joined to the imported location dataset so each site had a longitude and latitude. These cleaned files were then saved in the Data»Processed folder as csv.

Data Structure	Value
Variables	Site, Analyte, ppt, Total PFAS, Sum PFOA PFOS, short-chain, long-chain, latitude, longitude
Units	parts per trillion (ppt) or equivalent nanograms per liter (ng/L)
Range	2650 (POTW) and 552 (PWS)
Skew	14.28 (POTW) and 3.56 (PWS)
Kertosis	254.5 (POTW) and 15.2 (PWS)
Links to Data Sources	2018 PWS and 2019 POTWs

3 Exploratory Analysis

3.1 Unique analytes

45 unique analytes were found in POTW samples and 23 in PWS. The following two tables display the most abundant analytes found in each set of site samples.

Table 2: Top 10 Abundant Analytes in POTW

Analyte	Count
PFOA	74
PFHxA	73
PFOS	73
PFHpA	68
PFPeA	54
PFNA	45
PFDA	42
PFHxS	33
PFBS	26
PFBA	25

Table 3: Top 10 Abundant Analytes in PWS

Analyte	Count
PFPeA	17
PFHxA	15
PFHpA	14
PFBA	10
PFOS	8
PFOA	4
PFHxS	2
4.2	0
6.2	0
8.2	0

3.2 Number of samples with each analyte

The following histograms visualize the number of samples of each analyte. Same as above, the abundnace indicates which PFAS are most commonly present. This is relevant for regulation and possibly tracking contaminant sources.

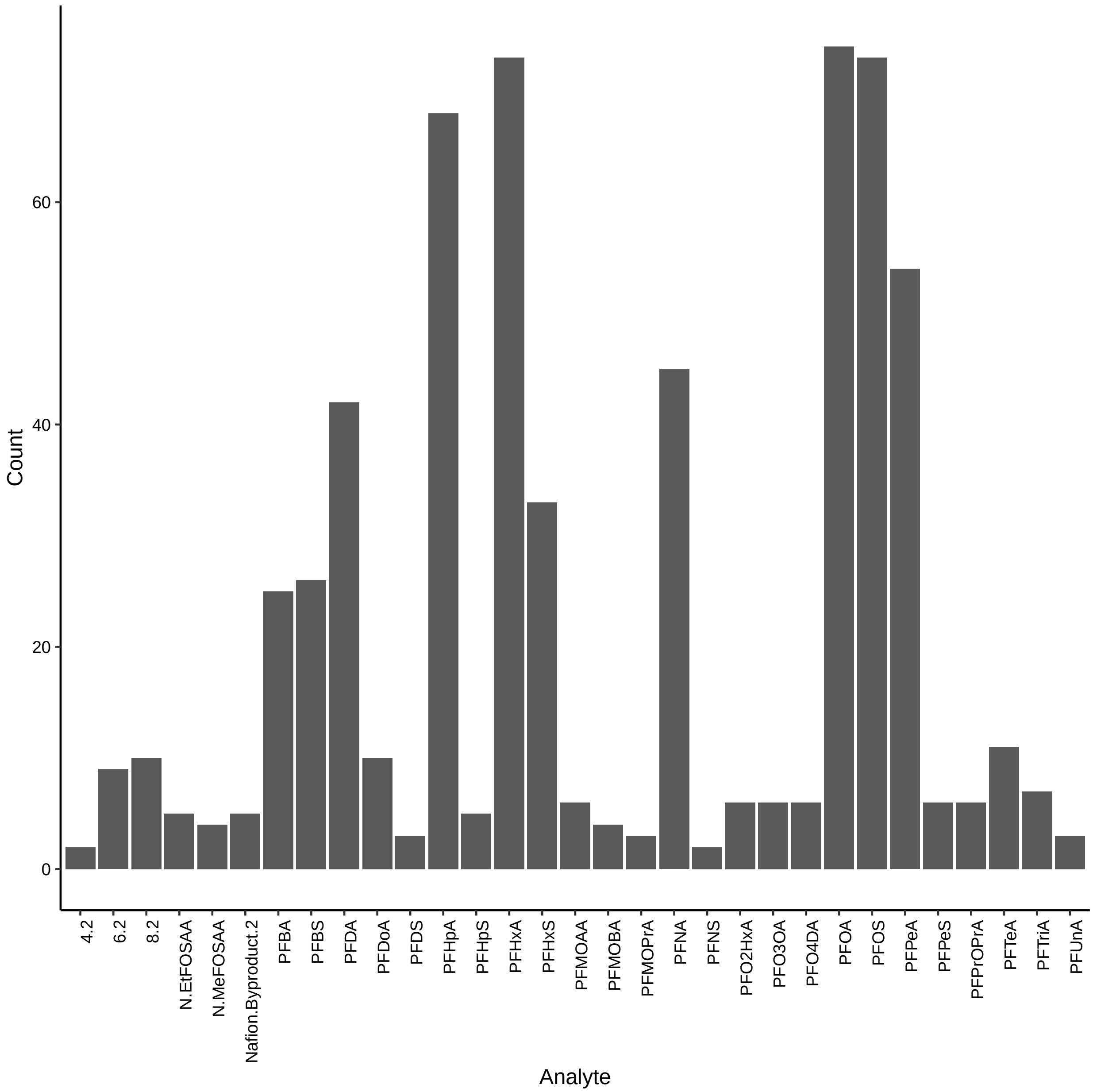


Figure 1: Number of Samples in POTW

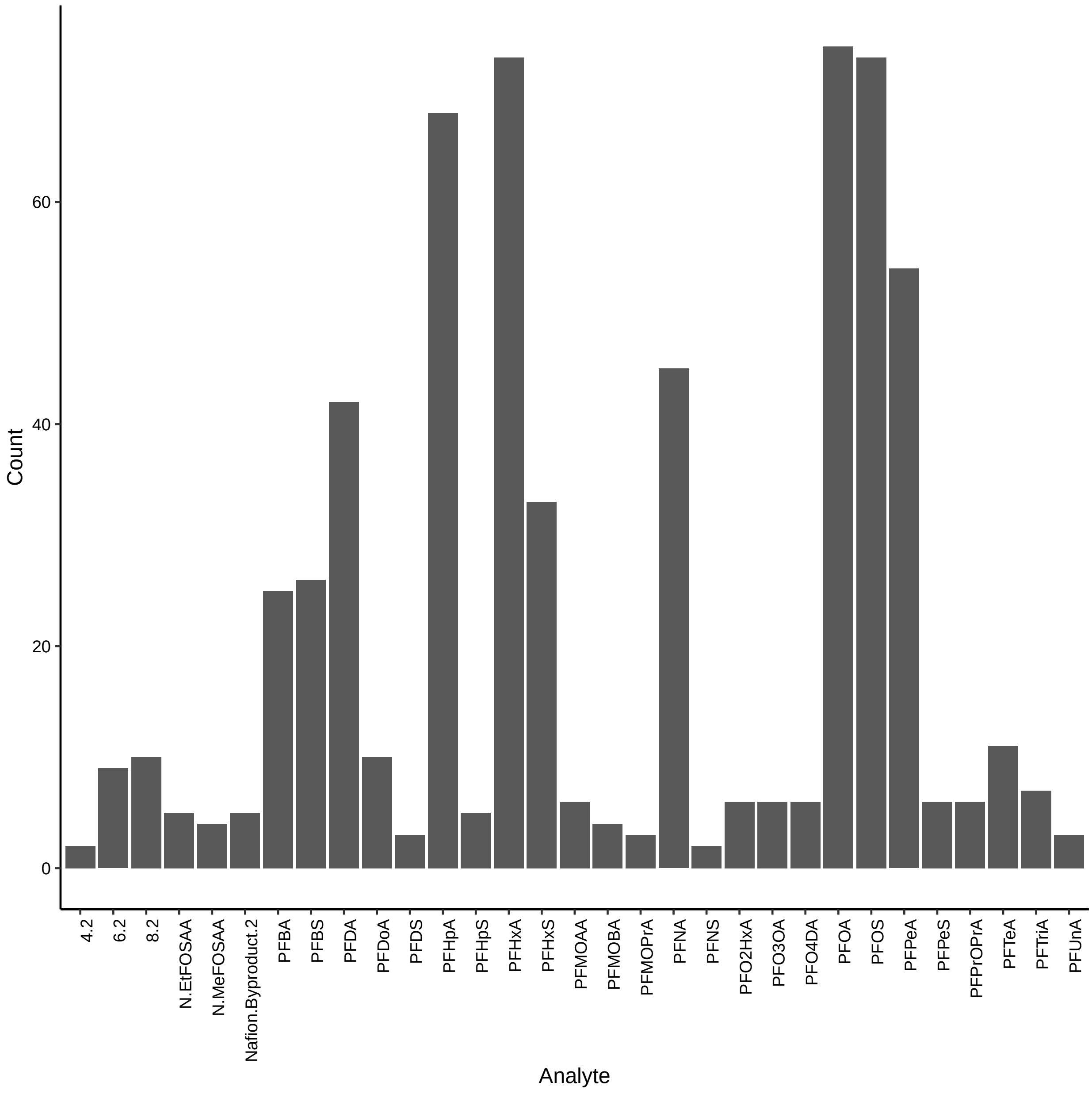


Figure 2: Number of Samples in PWS

3.3 Number of Select Analytes by Level

The following histograms visualize the number of samples of PFAS levels. The abundnace indicates which PFAS are most commonly present and which analytes have a high count at higher levels (ppt). This is relevant for regulation and possibly tracking contaminant sources.

3.4 Density ridge lines of select analytes

A closer look the count of analyte samples by value (ppt) indicates which analytes have frequently high values. This is valuable knowledge to begin analyzing whether high levels are an isolated event or consistently high.

3.5 Distribution of PFOA and PFOS samples

Violin plots illustrating the distribution of PFOA and PFOS in both sample site types indicate the quartiles of both analytes. This is important for understanding the PFAS levels where the bulk of samples occured.

3.6 PFOA/PFOS exceeding health advisory in POTW

These tables shows the three sites that had the sum of PFOA and PFOS over 70ppt (EPA's health advisory level) during the three-month testing samples.

Table 4: Sites Exceeding EPA Health Advisory Limit

Site	PFOA	PFOS	Sum	Sample.Date
City.of.Raeford	NA	124.0	124.0	2019-09-16
City.of.Raeford	NA	73.6	73.6	2019-08-05
City.of.Raeford	NA	NA	0.0	2019-07-15
East.Burlington.WWTP	10.8	12.5	23.3	2019-09-17
East.Burlington.WWTP	64.6	56.4	121.0	2019-08-06
East.Burlington.WWTP	73.0	49.8	122.8	2019-07-16
Sanford-Big.Buffalo.Creek.WWTP	11.0	1000.0	1011.0	2019-09-04
Sanford-Big.Buffalo.Creek.WWTP	11.5	40.8	52.3	2019-08-06
Sanford-Big.Buffalo.Creek.WWTP	11.6	46.3	57.9	2019-07-08

3.7 Maximum reporting limits per analyte

If an analyte was measured below a reporting limit (ppt), it was not reported. Particularly high reporting limits are concerning because it indicates that there are sites that may have high levels of PFAS that are overlooked. While there is no specific value that is "too" high, long-chain PFAS are a known health risk at lower levels (e.g., 8ppt). These tables indicate the maximum reporting limits that occured for each analyte sampled.

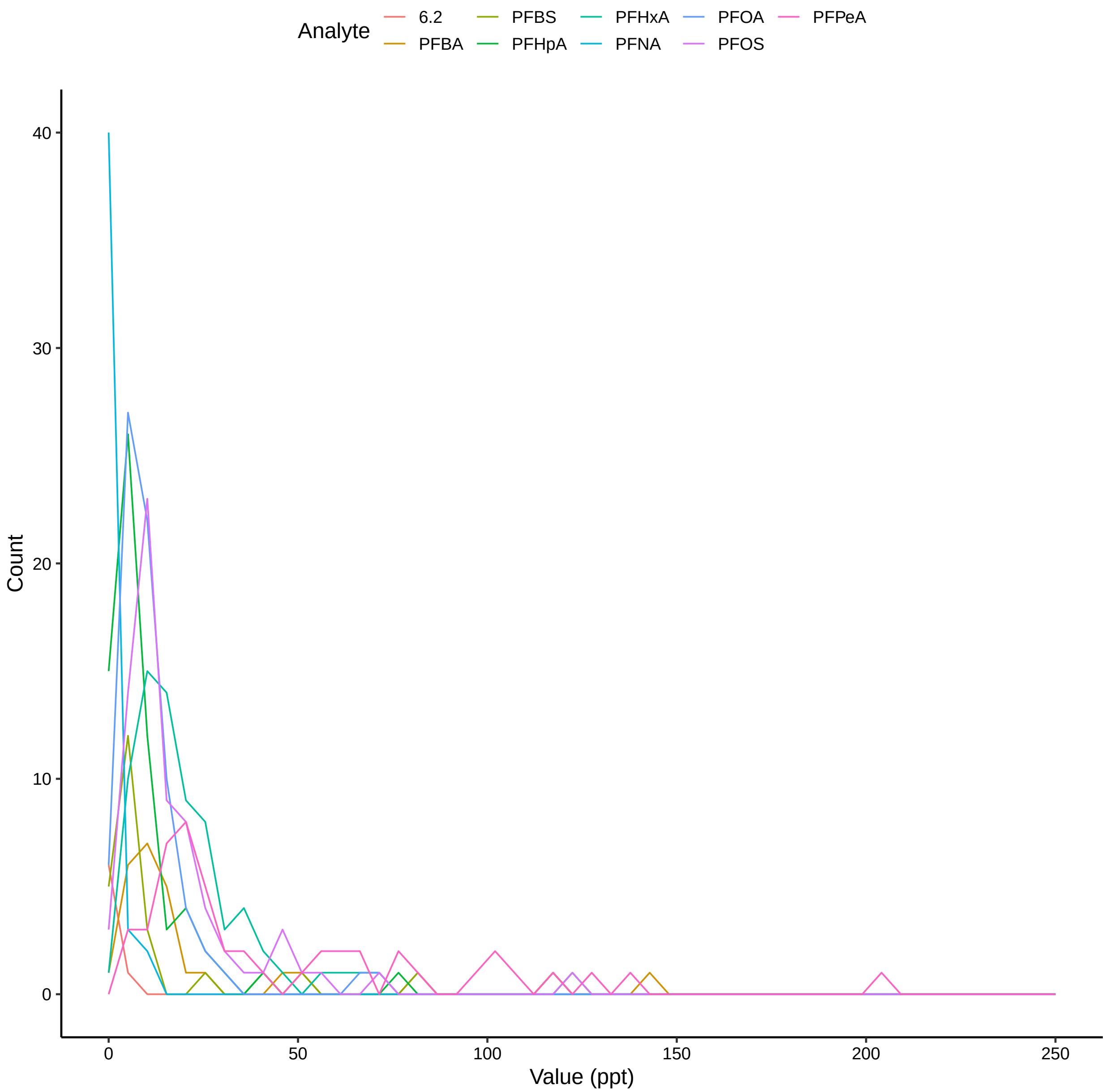


Figure 3: Number of Samples by Level in POTW

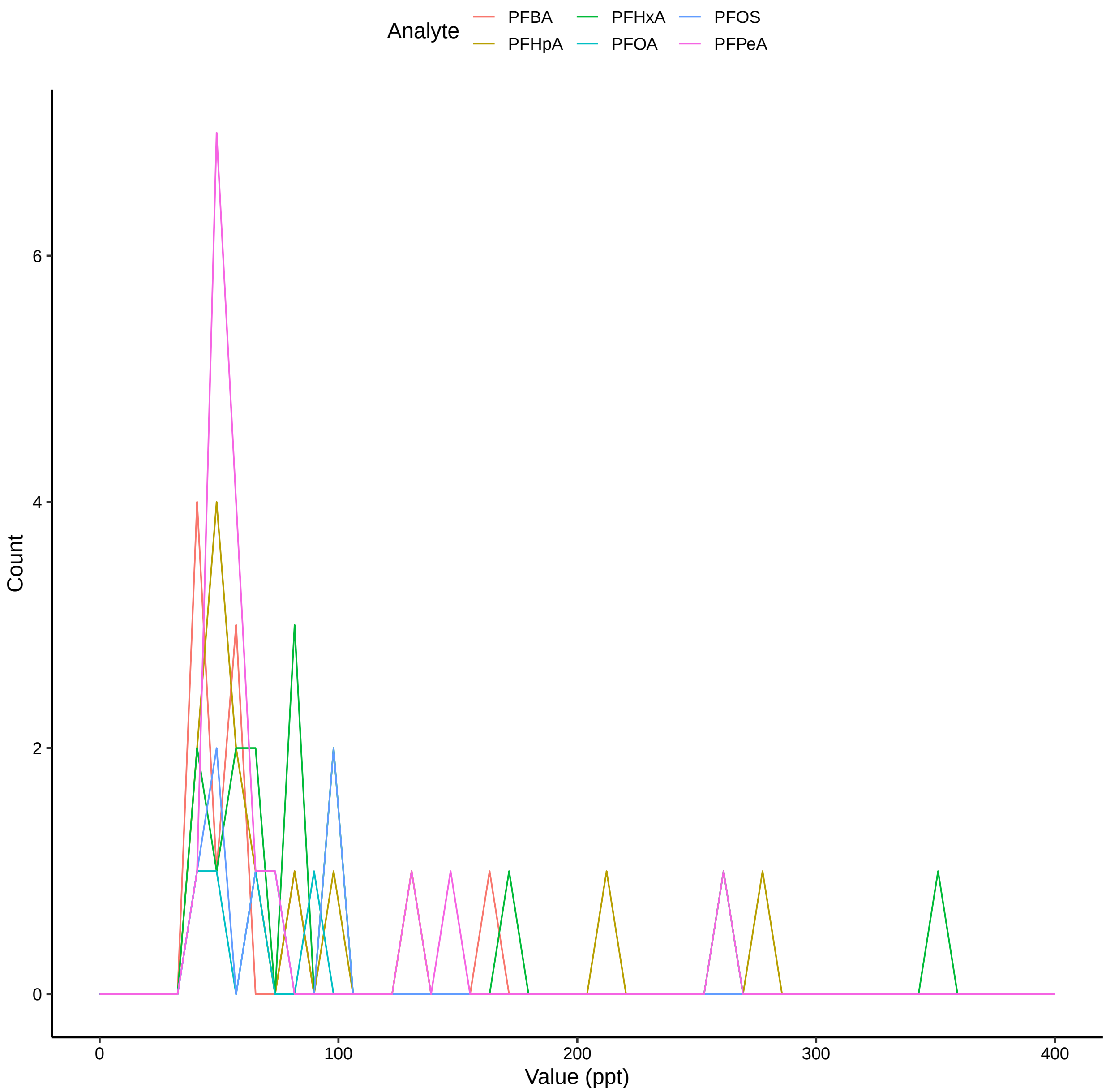


Figure 4: Number of Samples by Level in PWS

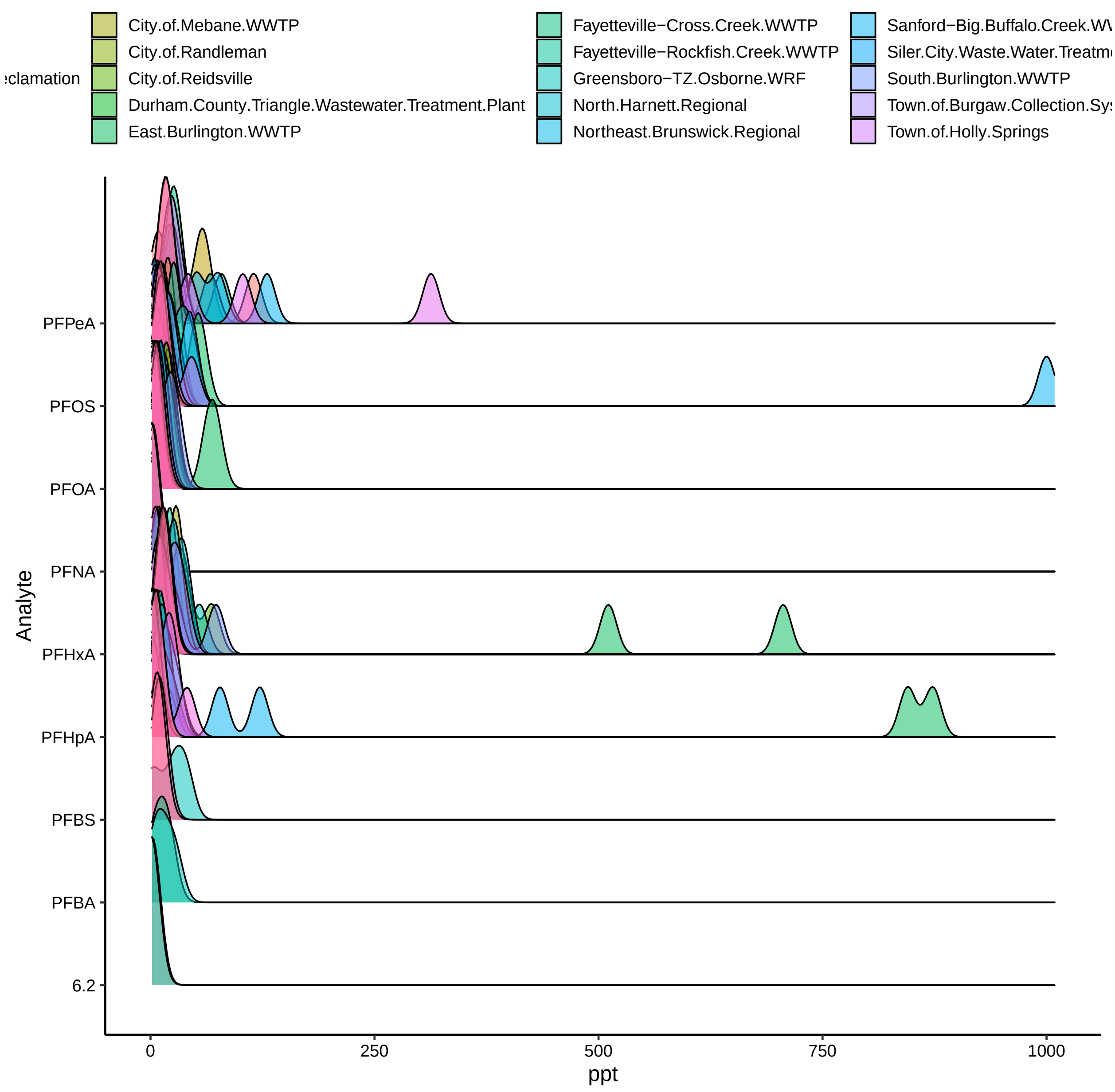


Figure 5: Density Ridges of Select Analytes by Level in POTW

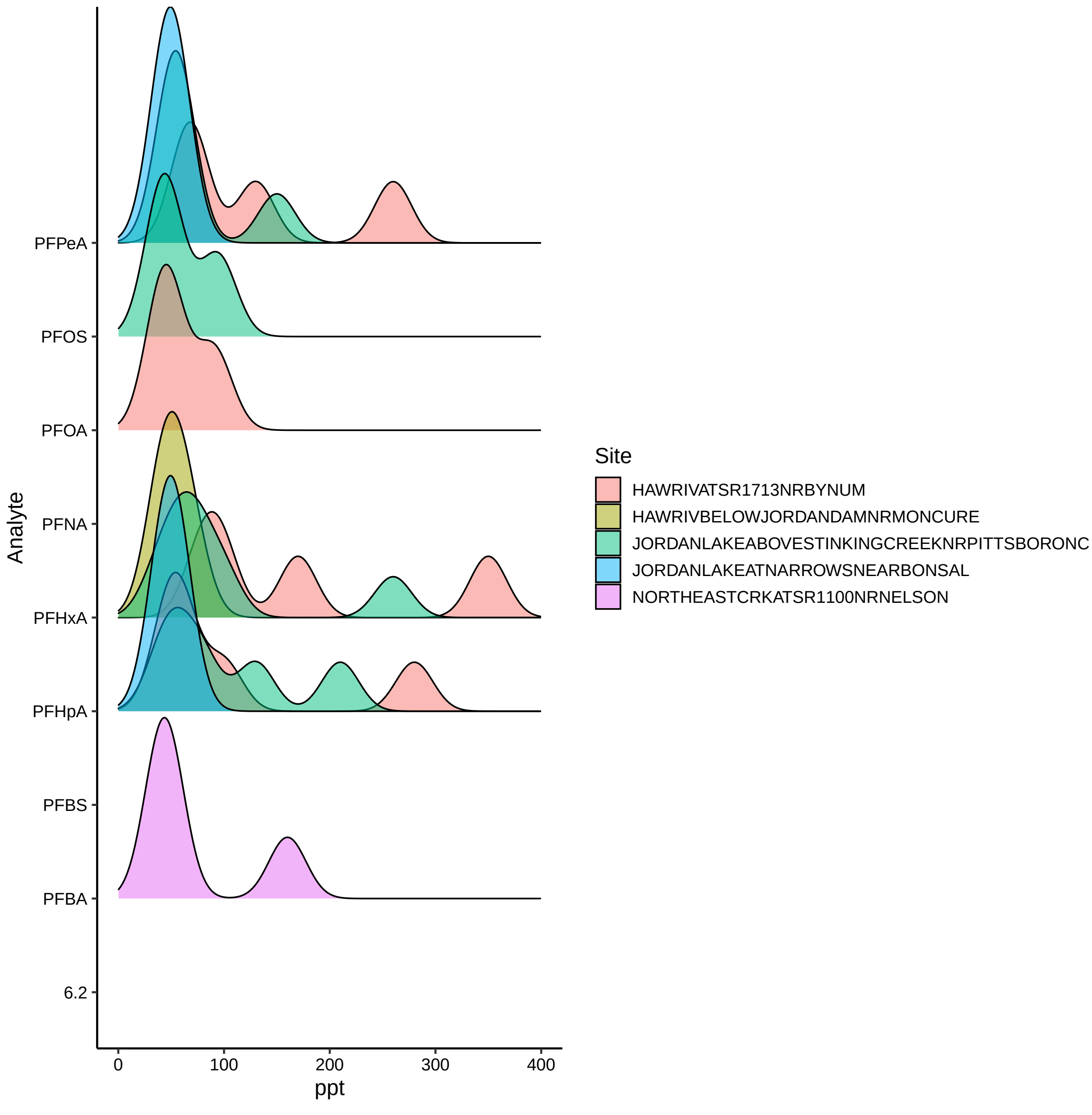


Figure 6: Density Ridges of Select Analytes by Level in PWS

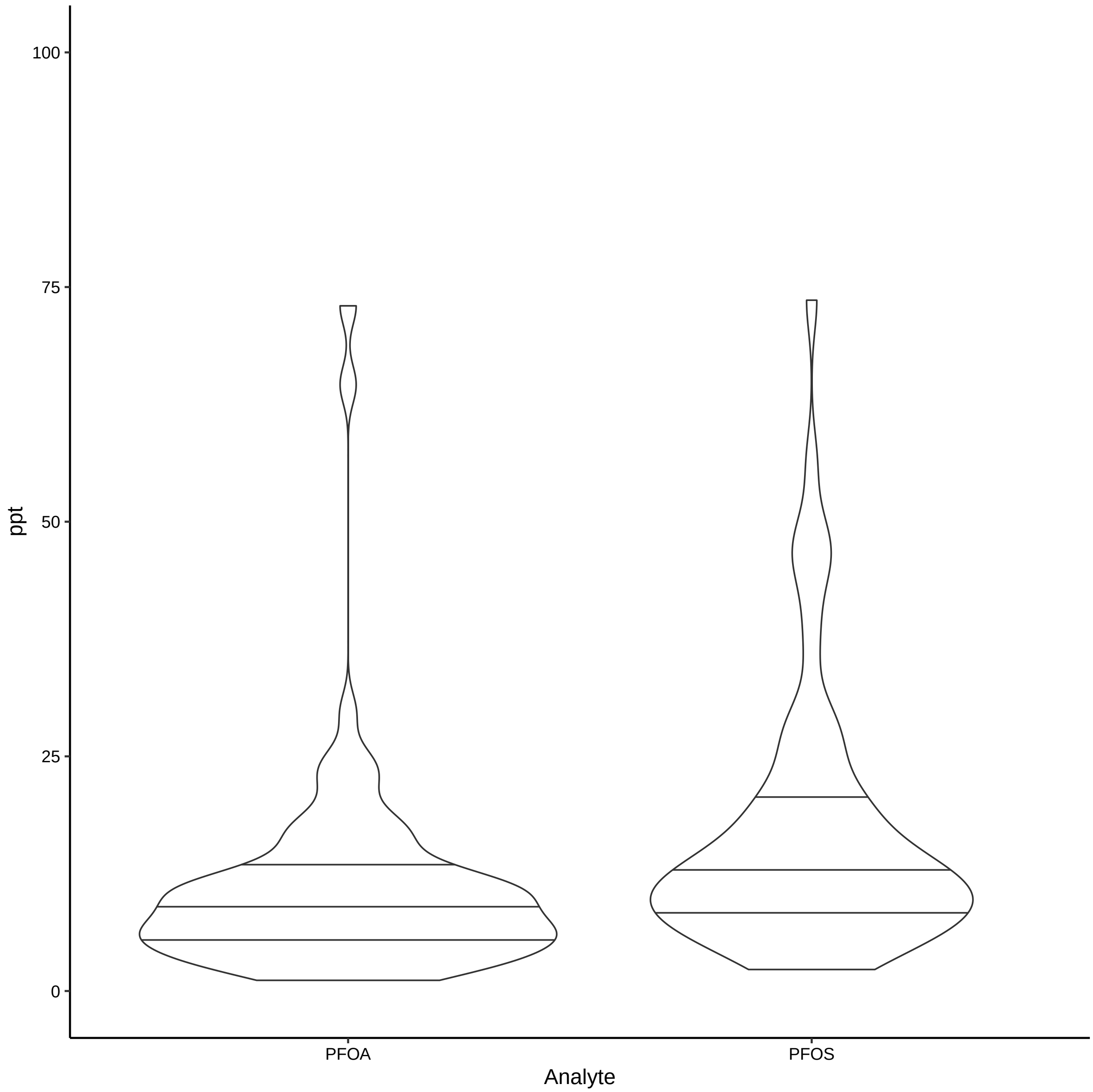


Figure 7: Violin Plots of PFOA and PFOS in POTW

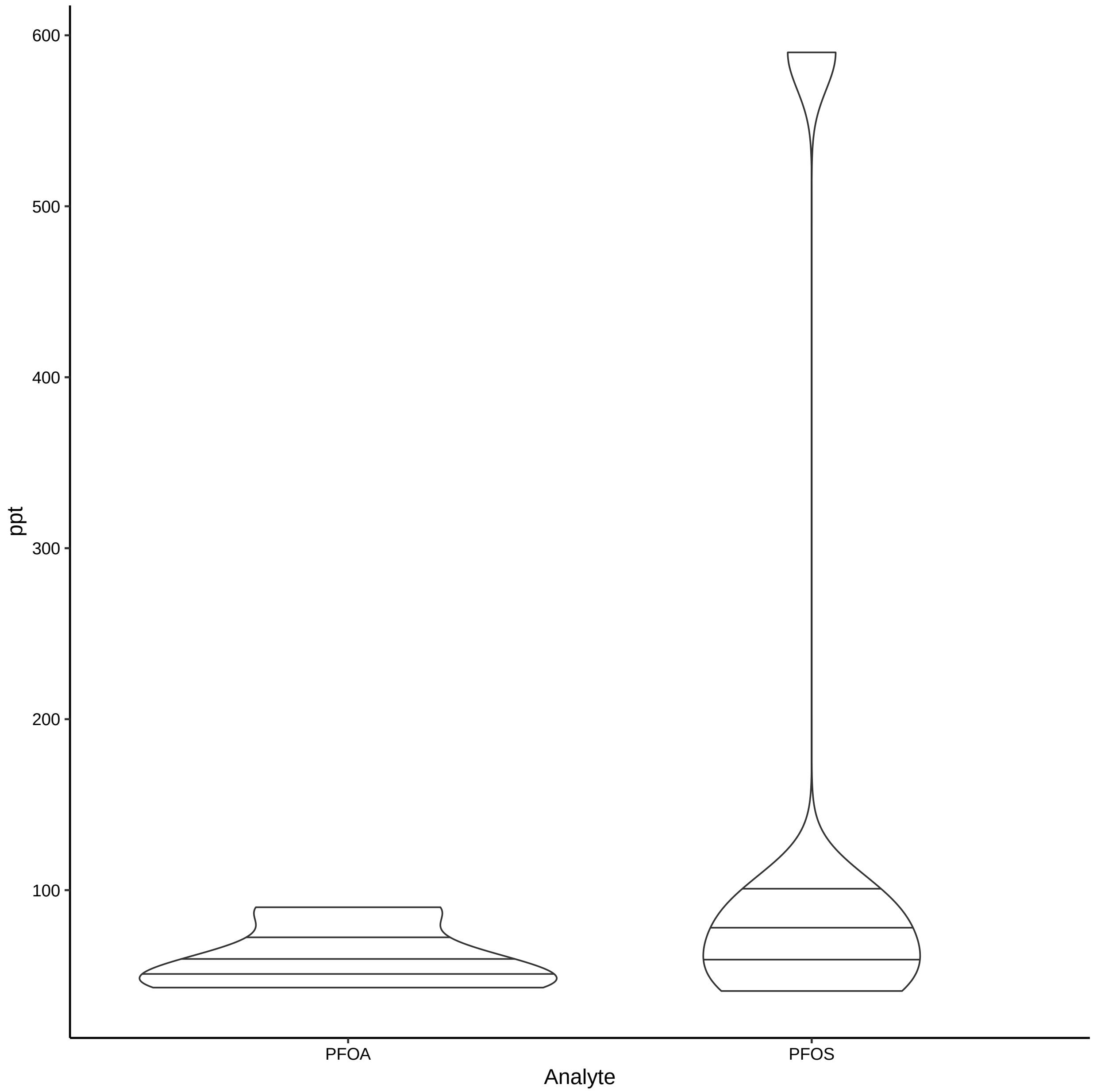


Figure 8: Violin Plots of PFOA and PFOS in PWS

Table 5: Max. Reporting Limits of Each Analyte in POTW

Analyte	Max. Reporting Limit
N.EtFOSAA	1900.00
N.MeFOSAA	1900.00
PFDoA	420.00
PFBA	380.00
PFDA	380.00
PFTeA	380.00
PFTriA	380.00
PFUnA	380.00
HFPO.DA	354.62
4.2	333.00
PFHxDA	210.00
PFNA	210.00
PFODA	210.00
PFDS	200.00
PFHpS	170.00
PFNS	170.00
PFOA	170.00
PFOSA	170.00
PFBS	120.00
PFHpA	120.00
PFHxA	120.00
PFHxS	120.00
PFOS	120.00
PFPeA	120.00
PFPeS	120.00
6.2	89.60
8.2	85.00
PFPrOPrA	21.00
N.MeFOSA	18.90
N.EtFOSA	18.70
N.EtFOSE	9.43
N.MeFOSE	9.43
Nafion.Byproduct.1	3.77
PFECA.G	3.77
PFMOPrA	3.77
10.2	3.62
Nafion.Byproduct.2	3.57
PFMOBA	3.57
PF3OUdS	3.55
PFMOAA	3.40

Analyte	Max. Reporting Limit
PFO2HxA	3.40
PFO3OA	3.40
PFO4DA	3.40
ADONA	1.89
PF3ONS	1.76

Table 6: Max. Reporting Limits of Each Analyte in PWS

Analyte	Max. Reporting Limit
6.2	310
PFUnA	310
NMeFOSAA	160
PFBA	160
PFDA	160
PFDoA	160
PFDS	160
PFNA	160
PFNS	160
PFTriA	160
8.2	83
4.2	82
PFHxA	82
PFOS	80
PFPeA	80
FOSA	42
HFPO.DA	42
PFBS	42
PFHpA	42
PFHpS	42
PFHxS	42
PFOA	42
PFPeS	42

4 Analysis

4.1 Question 1: Occurence

4.1.1 What are the most abundant analytes present?

The following two plots indicate the analytes present in POTW and PWS samples. First, the finding indicates that a wider variety of analytes were found in POTW compared to PWS, which may indicate additional or different contamination sources. It may also be an indicator of varioius PFAS precursors and or their degraded forms, which further research can focus on identifying patterns influencing those changes. Lastly, these graphs are the first clue that PFOA, PFOS, and GenX (the most commonly discussed, researched, and regulated analytes) may not be the most abundant PFAS we are exposed to via water.

4.1.2 Which sites have analytes exceeding health advisory levels?

Using the unique function allows a closer look at the number of sites with analytes above 30ppt and 70ppt (common health advisory or regulatory limits for PFOA and PFOS). In the POTW samples, there were 23 sites with analytes above 30ppt and 13 above 70ppt. In PWS samples, 11 sites were above 30ppt and 5 above 70ppt. Rather than comparing the two, these numbers indicate that a signficant number of sites had analytes at what could potentially be a concerning level, and further research is needed to determine if those specific analytes exceeding 30ppt and 70ppt pose a health risk. Certainly, they indicate the widespread presence of high levels of PFAS in North Carolina.

The following plots elaborate on this data, indicating which sites showed those respective levels. This information can be used to target subsequent sampling, depending on if the particular analytes pose a health risk to the communities nearby.

```
##  [1] City.of.Randleman
##  [2] Wallace.Regional.WWTP
##  [3] South.Burlington.WWTP
##  [4] Fayetteville-Cross.Creek.WWTP
##  [5] Greensboro-TZ.Osborne.WRF
##  [6] City.of.Raeford
##  [7] City.of.High.Point.Eastside.WWTP
##  [8] Town.of.Holly.Springs
##  [9] City.of.Reidsville
## [10] City.of.Mebane.WWTP
## [11] Durham.County.Triangle.Wastewater.Treatment.Plant
## [12] Wilmington-James.A.Loughlin(Northside.WWTP
## [13] Wilmington-M'Kean.Maffitt(Southside.WWTP
## [14] Cit.of.Clinton-Norman.H.Larkins.WWTP
## [15] City.of.Durham-South.Durham.Water.Reclamation
## [16] Sanford-Big.Buffalo.Creek.WWTP
## [17] North.Harnett.Regional
## [18] Town.of.Ramseur.WWTP
```

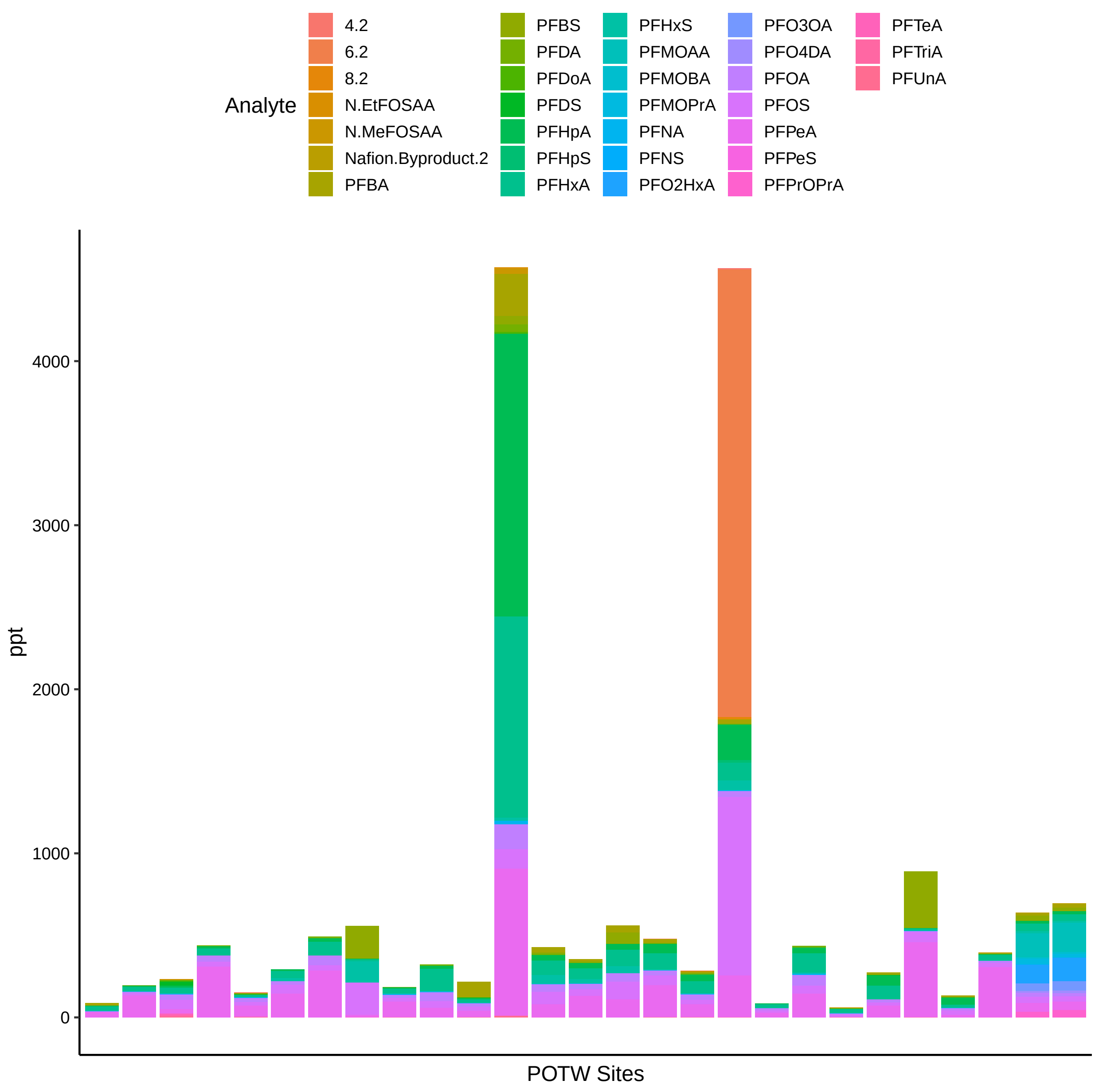


Figure 9: Analytes Present in POTW Sites

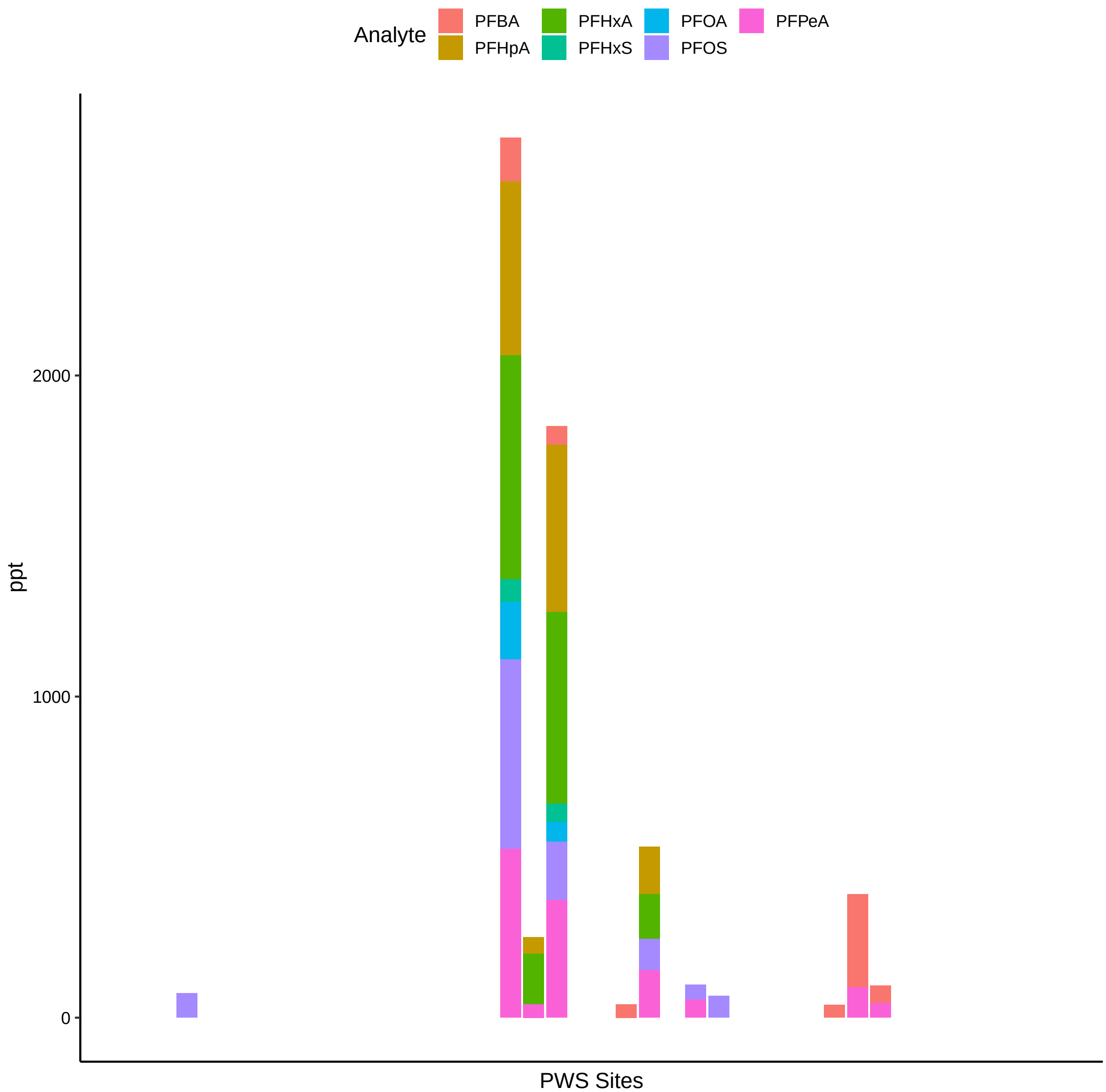


Figure 10: Analytes Present in PWS Sites

```
## [19] Fayetteville-Rockfish.Creek.WWTP
## [20] Town.of.Rose.Hill.WWTP
## [21] East.Burlington.WWTP
## [22] City.of.Graham
## [23] Northeast.Brunswick.Regional
## 27 Levels: Cary.Western.Wake.Regional.Water.Reclamation.Facility ...

##  [1] City.of.Randleman
##  [2] Wallace.Regional.WWTP
##  [3] South.Burlington.WWTP
##  [4] City.of.Raeford
##  [5] City.of.Mebane.WWTP
##  [6] Cit.of.Clinton-Norman.H.Larkins.WWTP
##  [7] City.of.Durham-South.Durham.Water.Reclamation
##  [8] Sanford-Big.Buffalo.Creek.WWTP
##  [9] North.Harnett.Regional
## [10] Town.of.Ramseur.WWTP
## [11] Fayetteville-Rockfish.Creek.WWTP
## [12] East.Burlington.WWTP
## [13] Northeast.Brunswick.Regional
## 27 Levels: Cary.Western.Wake.Regional.Water.Reclamation.Facility ...

##  [1] HAWRIVATSR1713NRBYNUM
##  [2] NEWHOPECRKATSR1107NRBLANDS
##  [3] NORTHEASTCRKATSR1100NRNELSON
##  [4] NORTHEASTCRKATSR1731OKELLYCHURCHRDNRDURHAM
##  [5] HAWRIVBELOWJORDANDAMNRMONCURE
##  [6] LakeBrandtatDamnearHillsdale
##  [7] JORDANLAKEABOVESTINKINGCREEKNRPITTSBORONC
##  [8] JORDANLAKEATMOUTHOFWHITEOAKCREEKNRSEAFORTH
##  [9] JORDANLAKEATNARROWSNEARBONSAL
## [10] CaneCreekReservoiratDamnrOaks,NC
## [11] KnapofReedsCreek
## 44 Levels: BlowingRockTownPond BooneASULake ... UniversityLakeatDamnrChapelHill,NC

## [1] HAWRIVATSR1713NRBYNUM
## [2] NORTHEASTCRKATSR1100NRNELSON
## [3] JORDANLAKEABOVESTINKINGCREEKNRPITTSBORONC
## [4] JORDANLAKEATNARROWSNEARBONSAL
## [5] CaneCreekReservoiratDamnrOaks,NC
## 44 Levels: BlowingRockTownPond BooneASULake ... UniversityLakeatDamnrChapelHill,NC
```

4.1.3 What chain lengths are most abundant and present?

PFAS have characteristically strong carbon bonds, and depending on their chain length, are categorized as either short- or long-chained PFAS. Long-chain PFAS are most researched

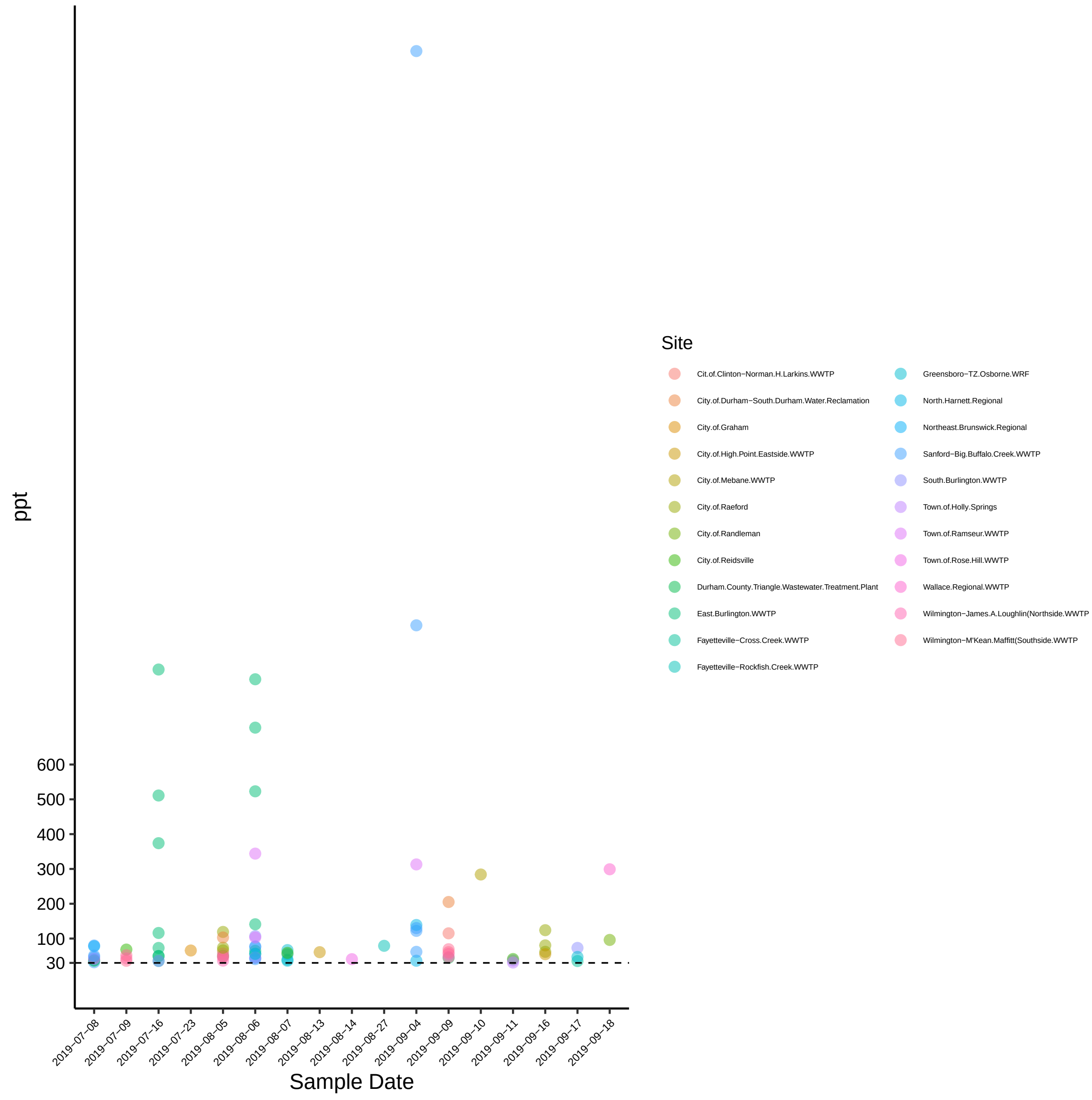


Figure 11: POTW Sites with Analytes Over 30ppt

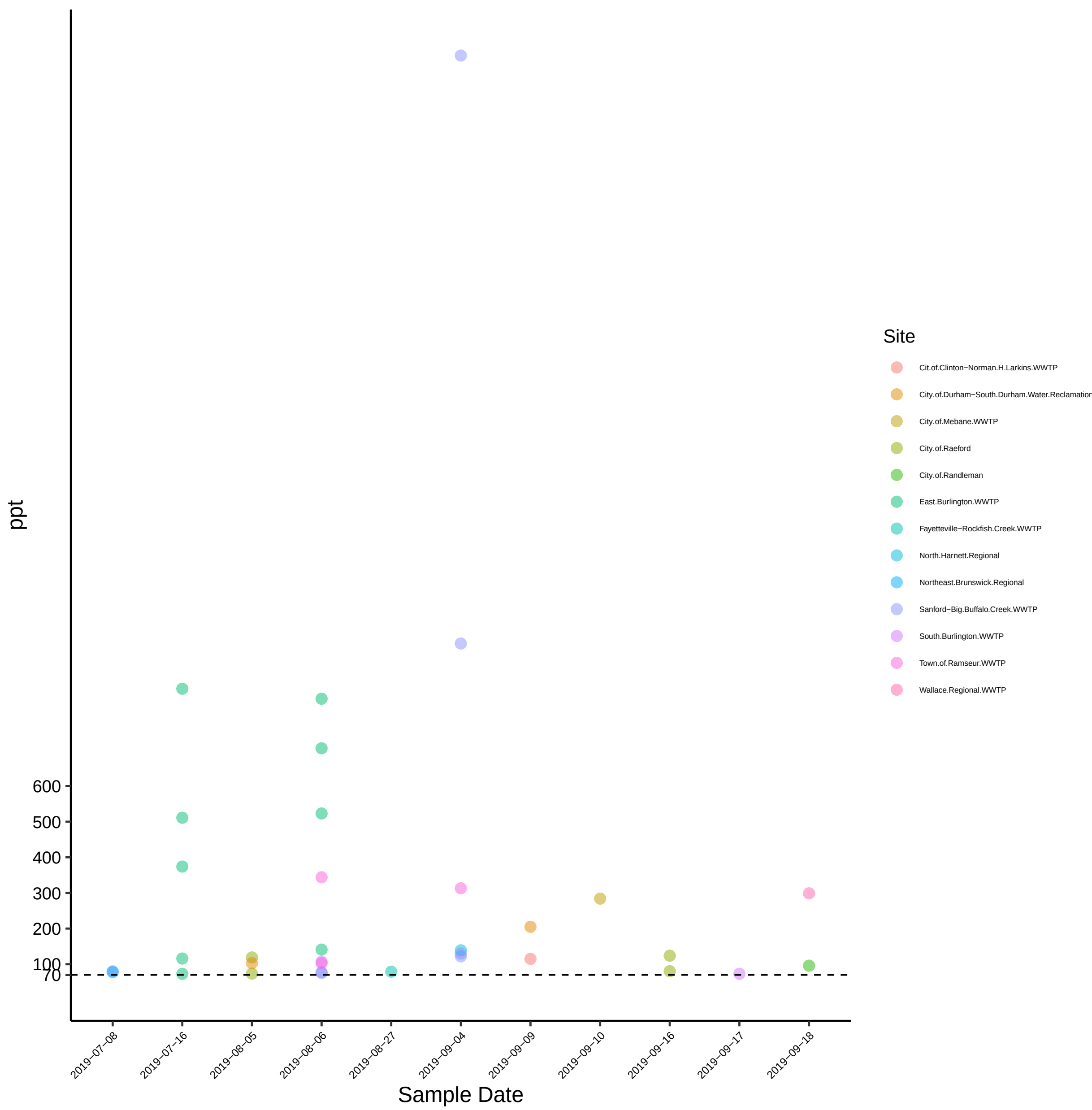


Figure 12: POTW Sites with Analytes Over 70ppt

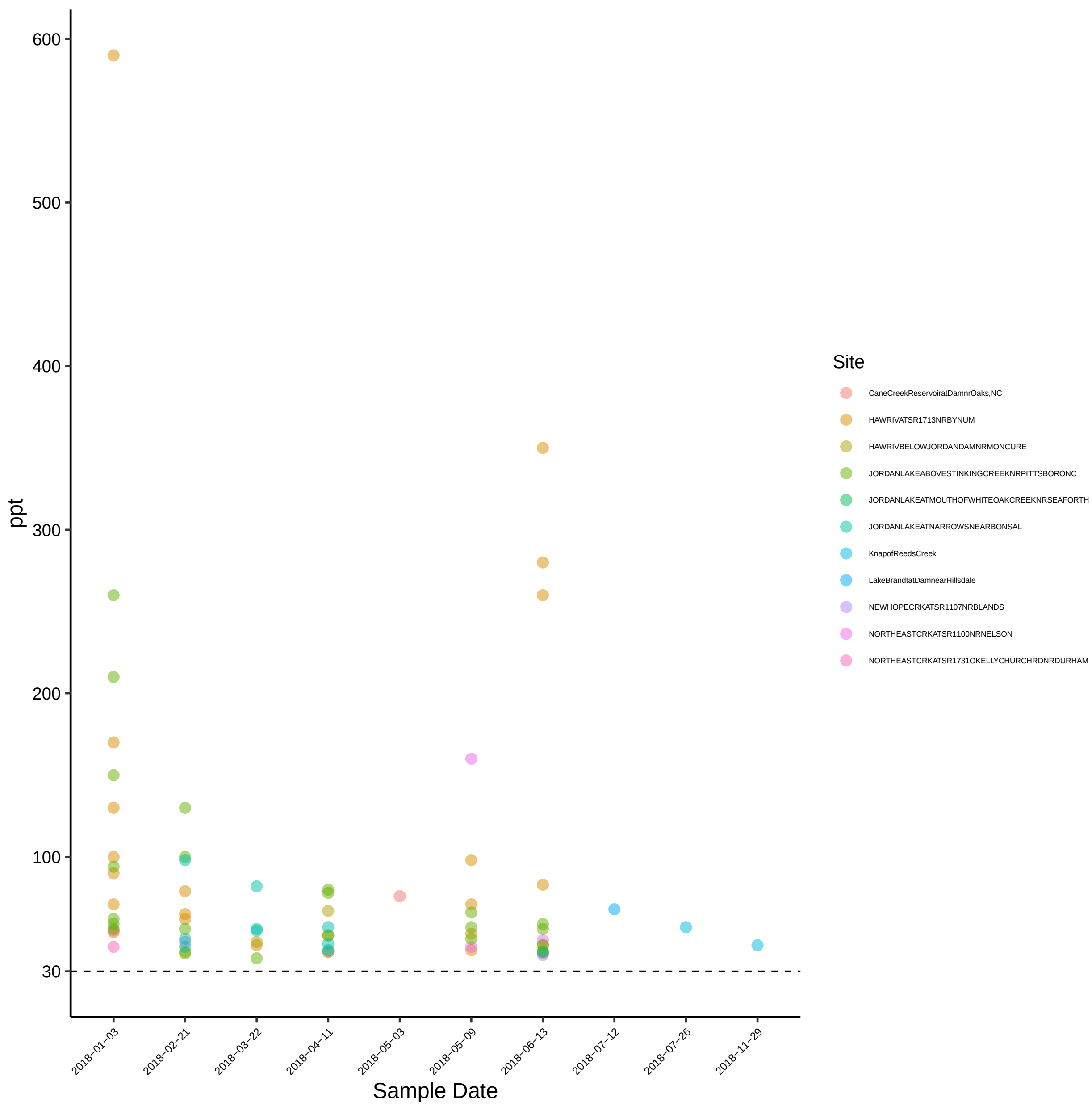


Figure 13: PWS Sites with Analytes Over 30ppt

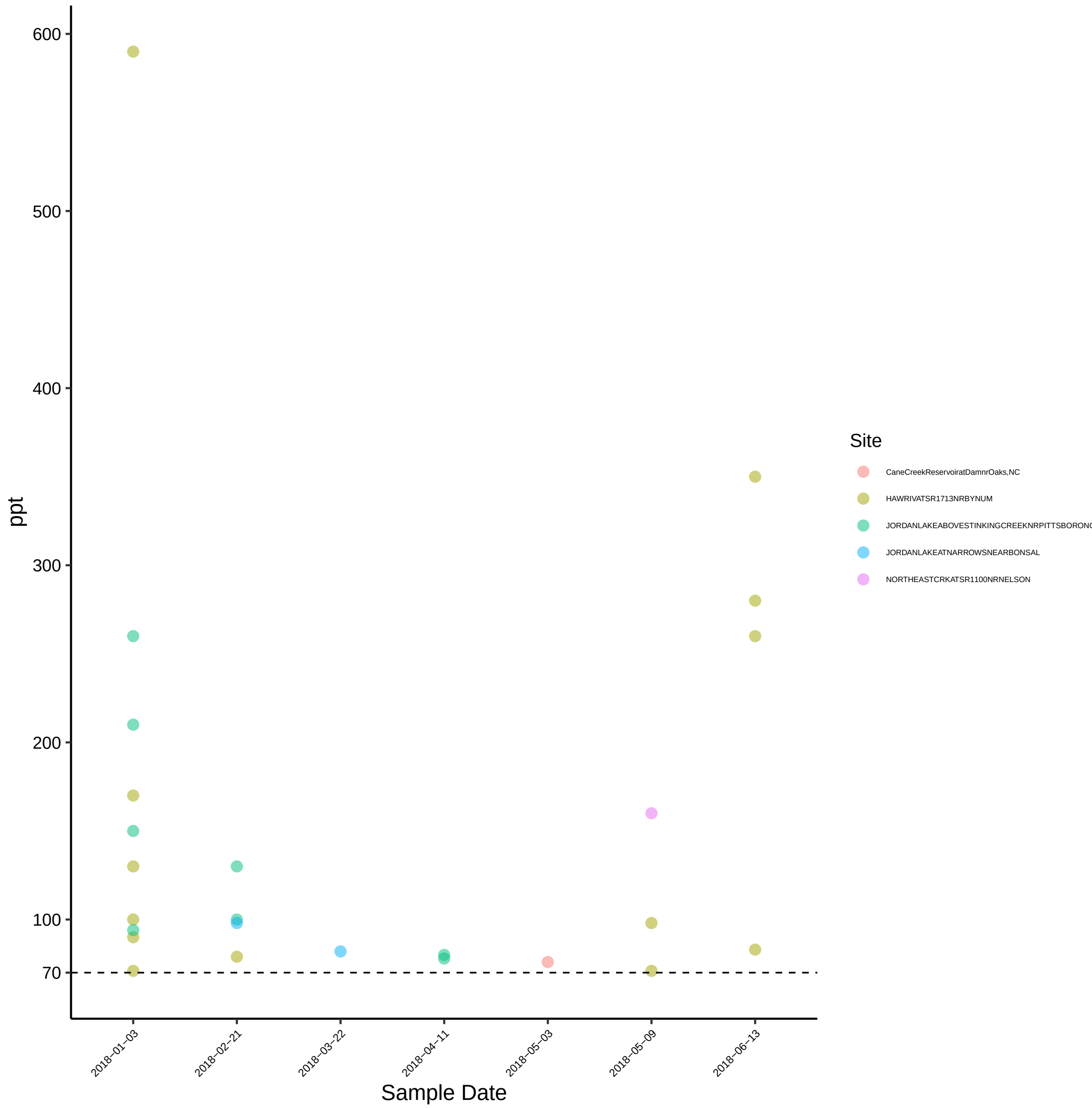


Figure 14: PWS Sites with Analytes Over 70ppt

(e.g., PFOA, PFOS) and take a long time to degrade. Furthermore, certain remediation treatments (such as the common and most affordable granular activated carbon (GAC)) are only effective at removing long-chain PFAS.

The following graphs indicate the proportion of total PFAS by chain length. This information should inform remediation and treatment upgrades as well as be interpolated to identify risks (for example, sites with predominantly short-chained PFAS but with GAC treatment may falsely assume it is adequately treating the risk to its community).

4.1.4 What portion of total PFAS are PFOA and PFOS?

This graph of POTW sites further illustrate the concern that the public is over-focused on PFOA and PFOS and is overlooking potential dangers from high total PFAS. Specifically, sites such as Sanford-Big Buffalo Creek get a lot of public attention because it had a high level of PFOA/PFOS but sites like the Town of Ramsur, Rose Hill, and Wallace Regional WWTPs had total PFAS above 250ppt and are infrequently mentioned.

4.2 Question 2: Distribution

4.2.1 Are there bodies of water with concerningly high levels of PFAS that aren't commonly discussed?

Expanding upon Question 1, graphing the top 12 highest total PFAS sites is the first step for researchers and regulators to focus further investigatory sampling. As mentioned before, in addition to the commonly publicized Burlington and Sandford-Big Buffalo Creek WWTPs, the Town of Ramseur, City of Mebane, Wallace Regional City of Raeford, Greensboro, Wilmington (North and Southside), North Harnett Regional, and City of Durham/South Durham Water Reclamation sites have surprisingly high total PFAS.

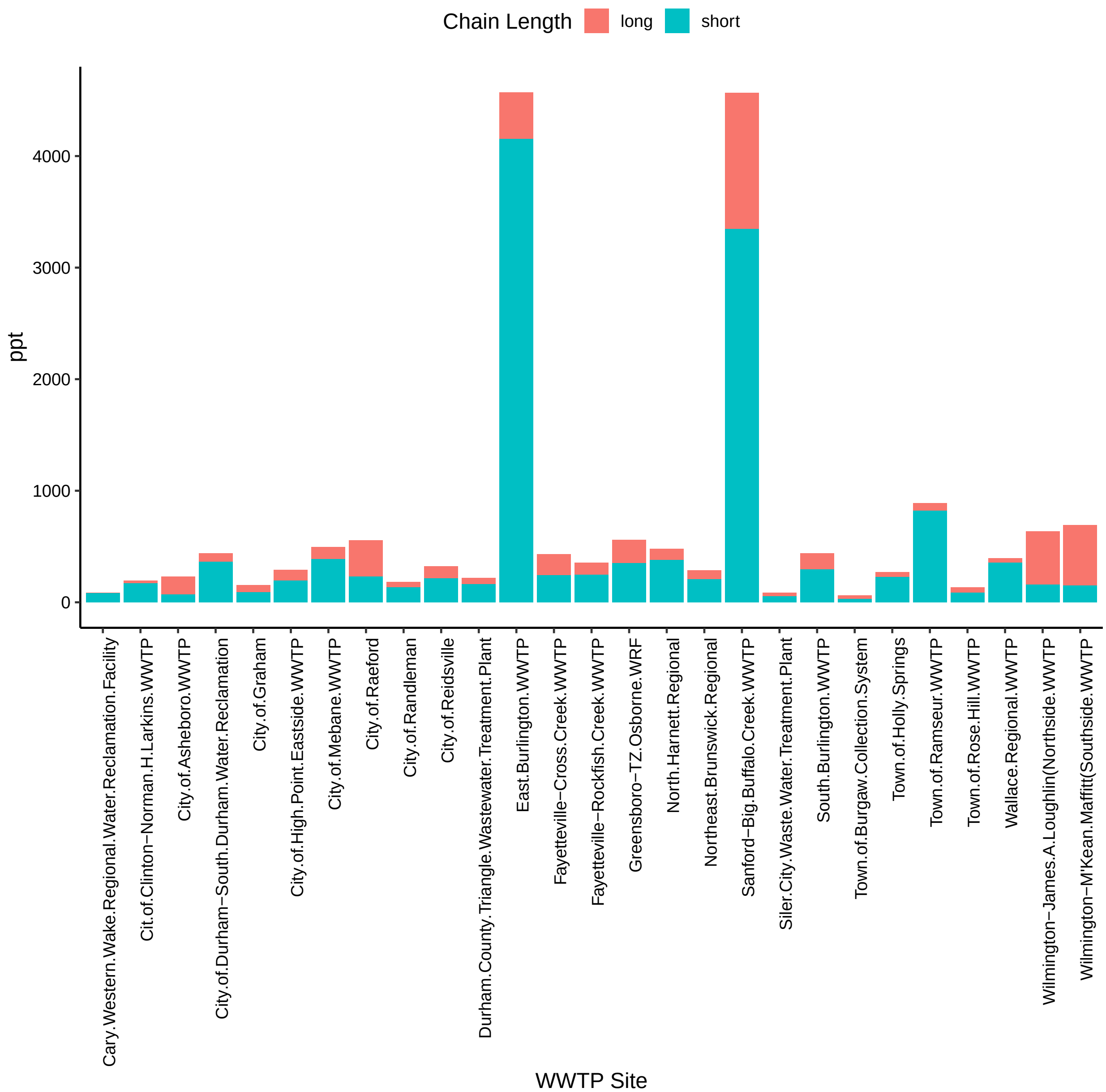


Figure 15: Total PFAS by Chain Length in POTW

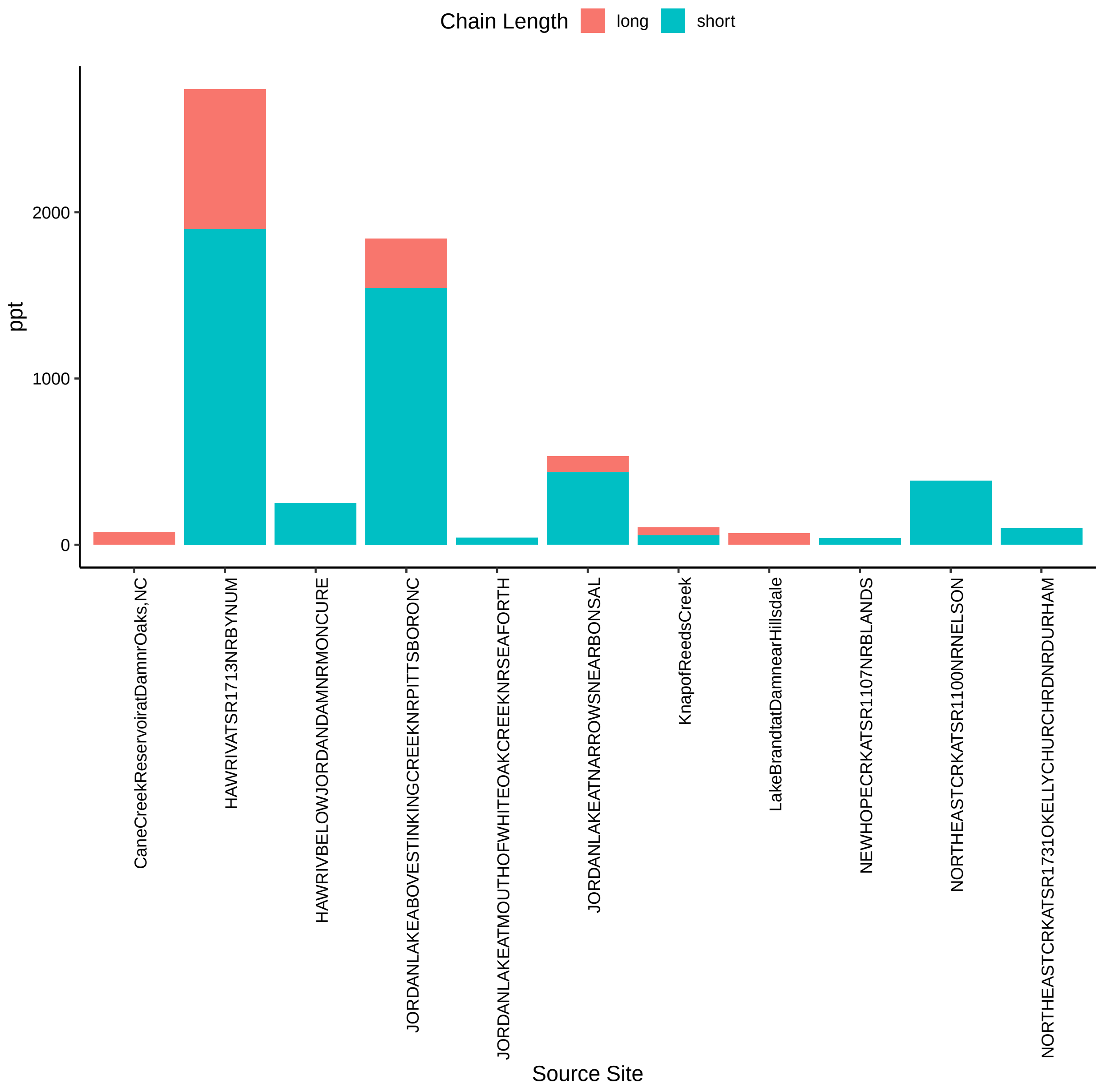


Figure 16: Total PFAS by Chain Length in PWS

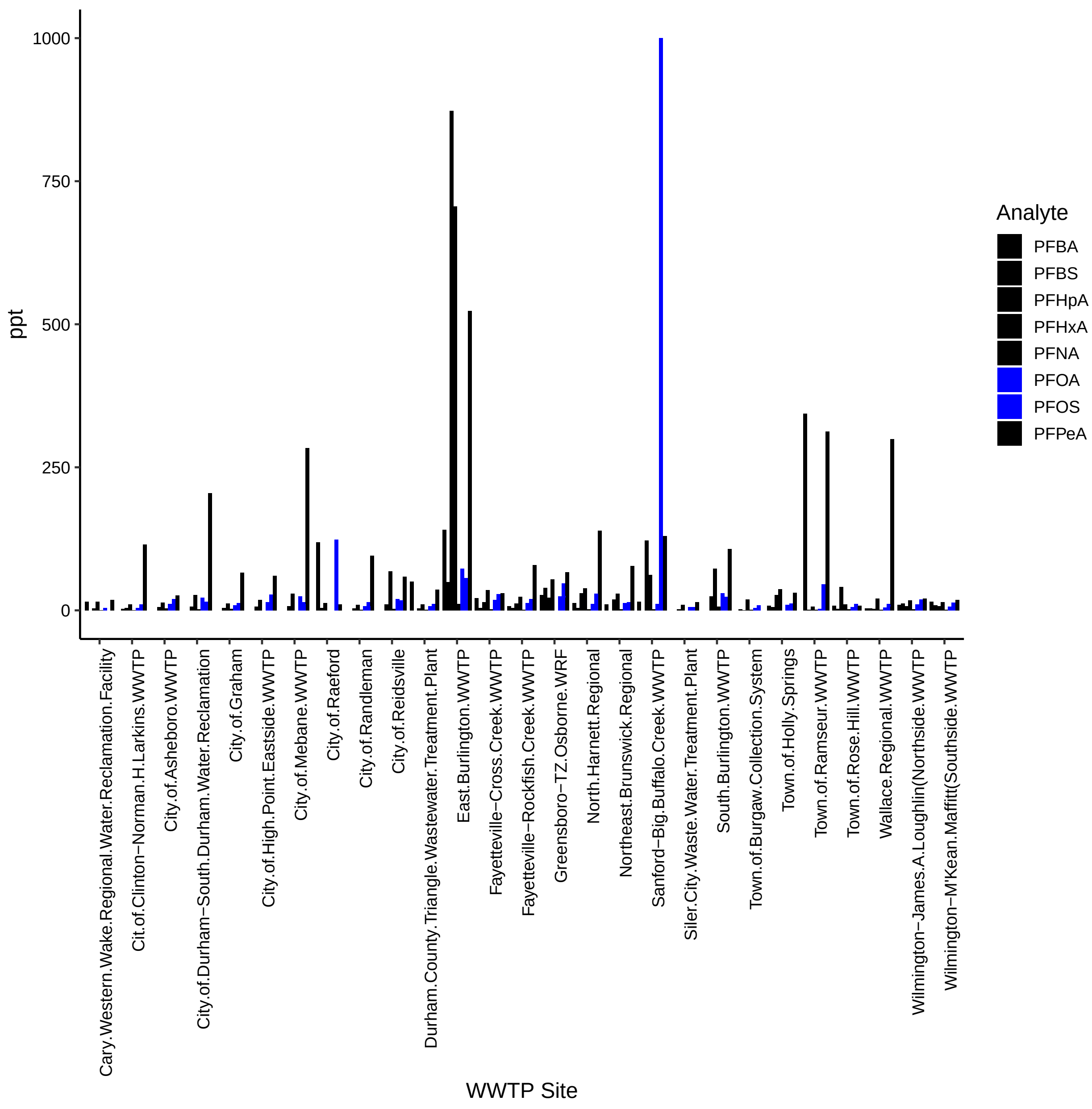
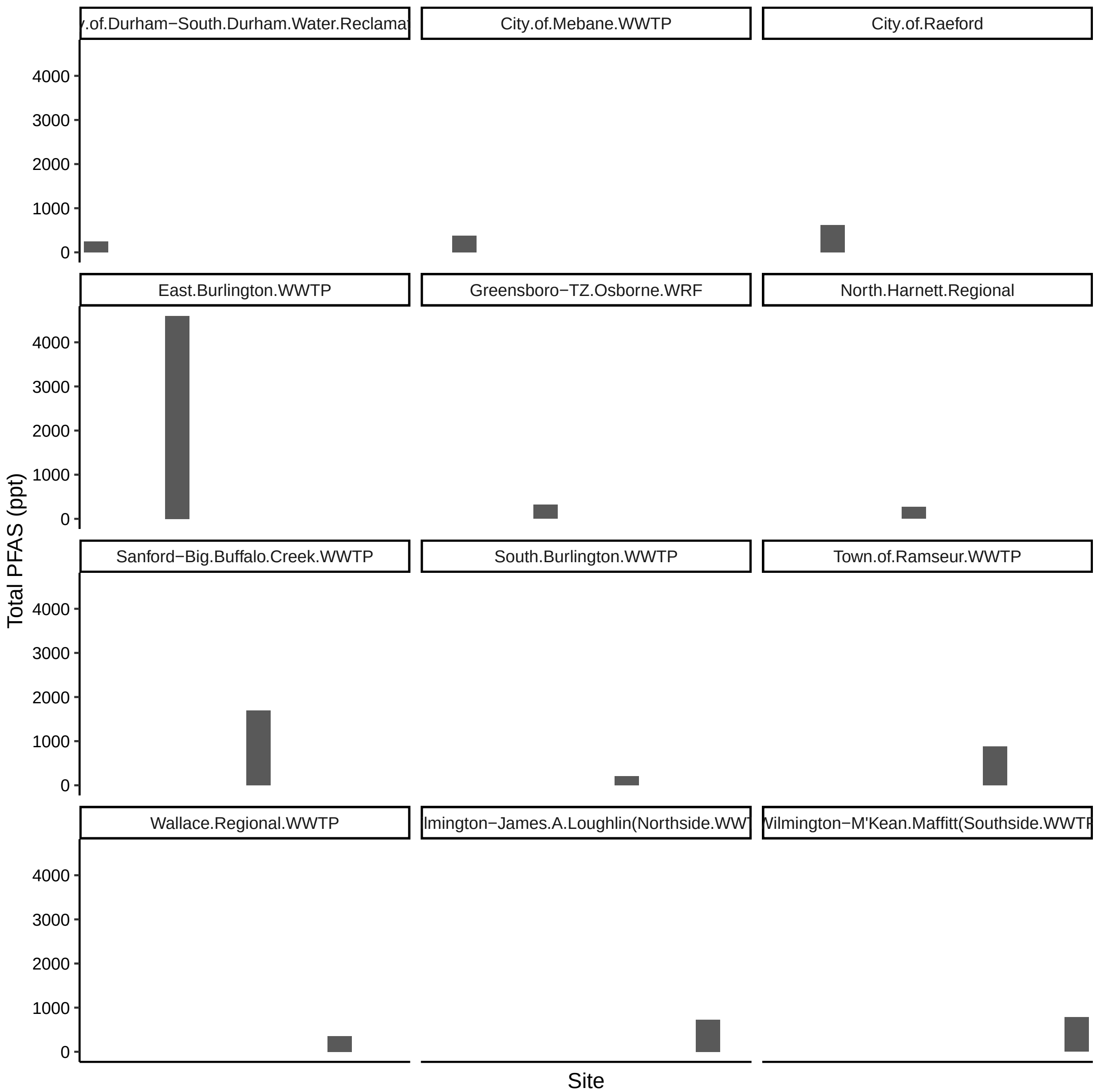


Figure 17: Proportion of PFOA and PFOS in Total PFAS at POTW Sites



4.2.2 How does total PFAS change over time?

Only the POTW dataset provided enough data for a temporal analysis, but it is ill-advised to make broad conclusions based on a three-month testing period. The first graph seems to indicate that, although there are isolated spikes, total PFAS remained consistently high and possibly increasing. A closer look (second graph removing outlier) illustrates that there was a definitive spike across multiple sites in August. Further research and more extensive sampling is needed to identify whether this is a temporal/seasonal trend, or if there were source-specific variables that impacted these levels.

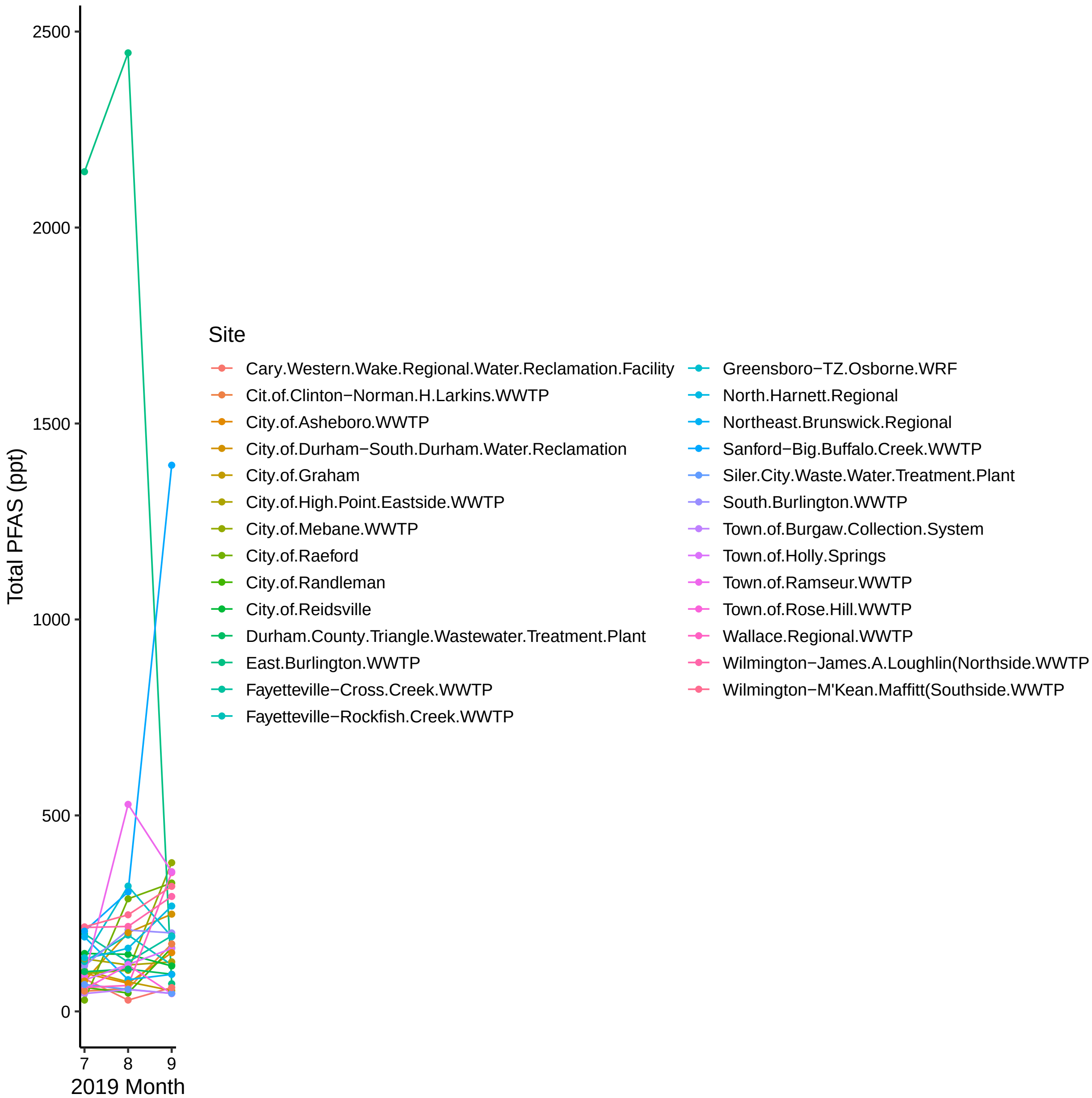


Figure 18: Total PFAS Change Over Time in POTW

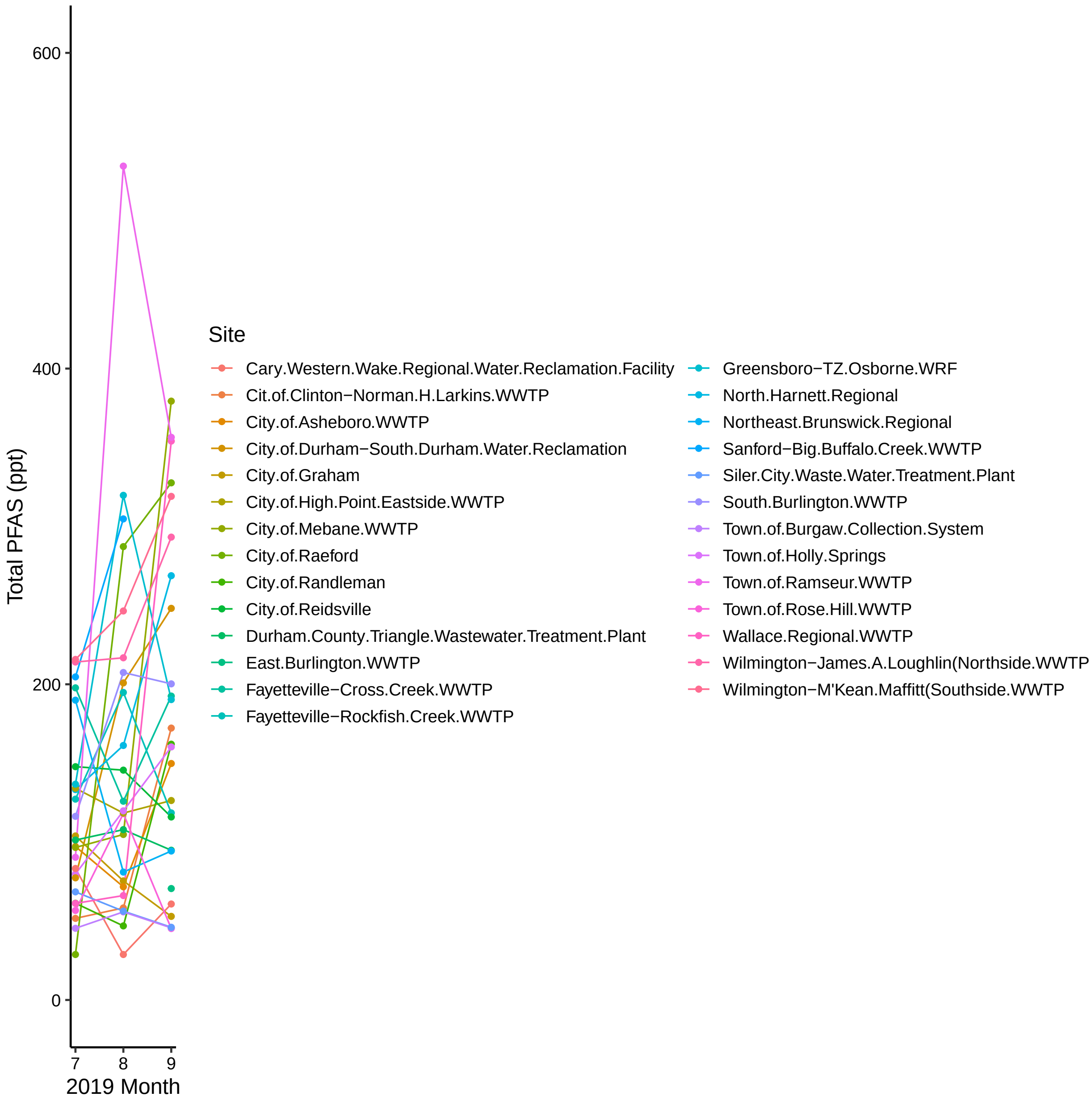


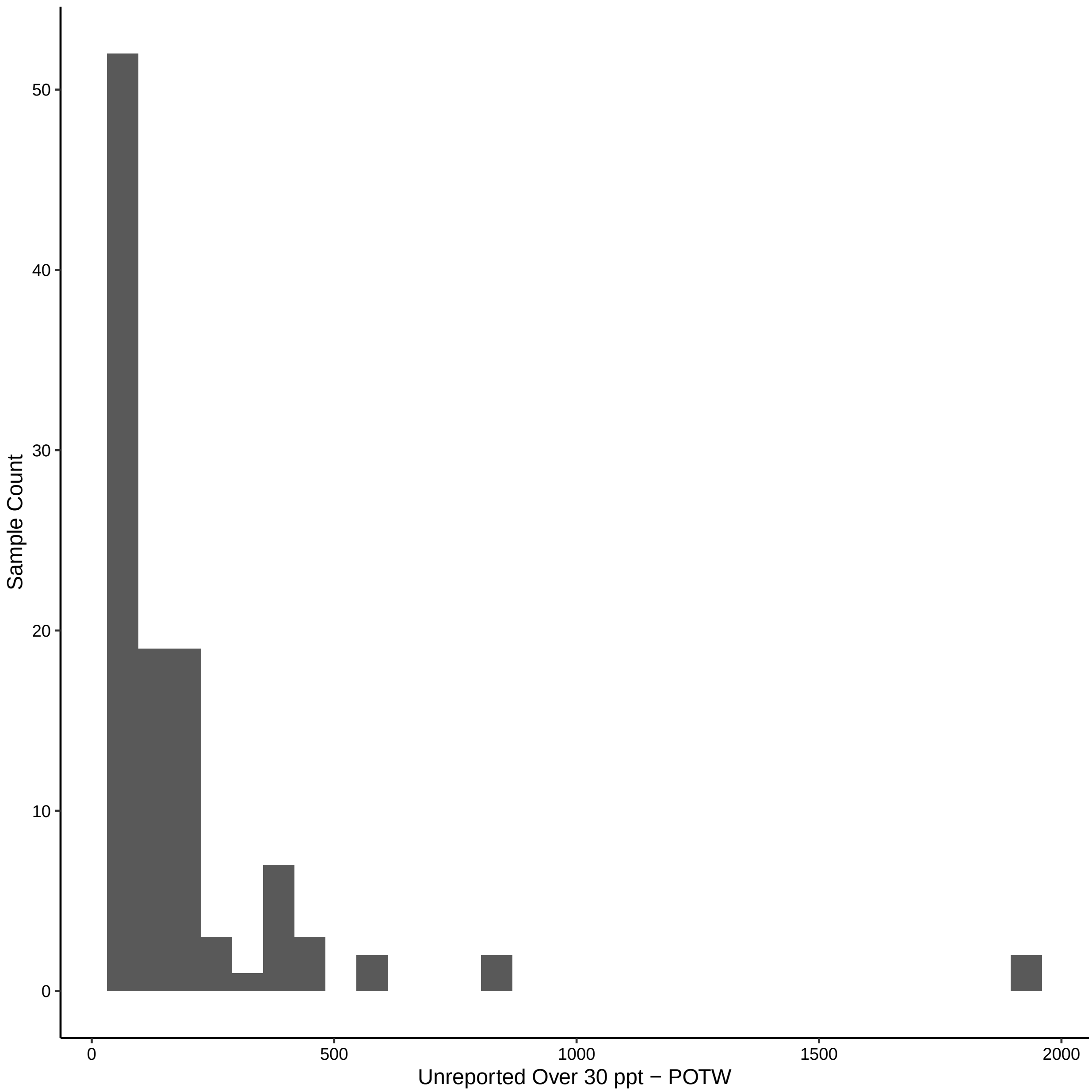
Figure 19: Total PFAS Change Over Time in Select POTW Sites

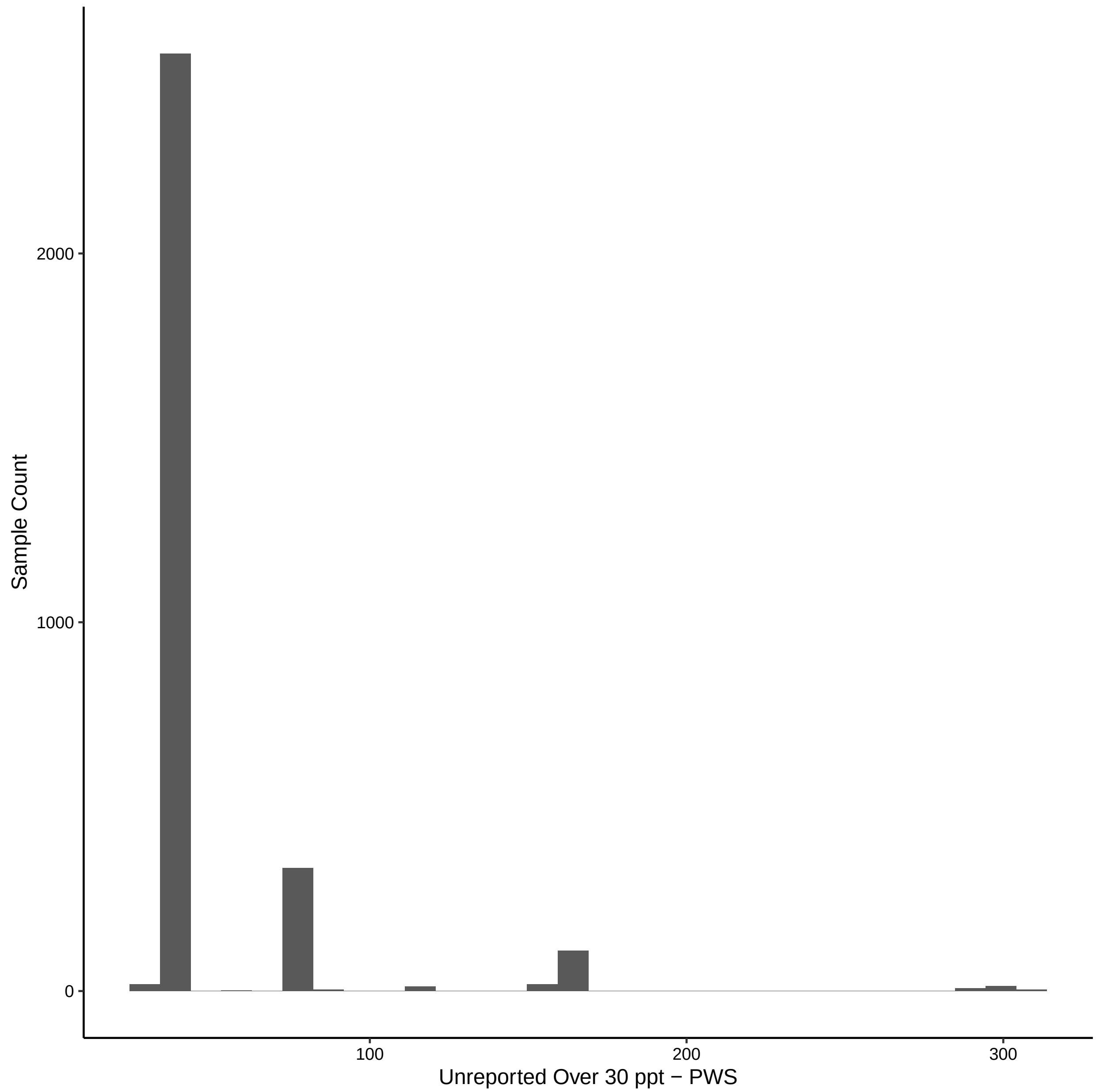
4.3 Question 3: Further Investigation

4.3.1 How many samples had reporting limits above health advisory levels?

As demonstrated earlier in the exploratory section, both datasets indicated reporting levels and lab qualifiers with relevant results. Concerningly, many samples had high reporting limits meaning that many samples did not have a reported value.

Using 30ppt as a benchmark for what a potentially conceringly high reporting level is, the following histogram illustrates the number of samples that had reporting limits above 30ppt. From these results, it can be inferred that a signficant number of samples had unreported analytes above 30ppt, meaning there are potentially many sites with concerning levels of PFAS analytes that are being overlooked and possibly mis-identified as 'safe'. Note that PWS had an overwhelmingly large number of unreported analyte samples due to these reporting limits.


Sample Count
50
40
30
20
10
0
0
500
1000
1500
2000
Unreported Over 30 ppt − POTW



4.3.2 What percentage of total unreported results were above health advisory limits?

Looking closer at the previous results, the following percentages clue researchers and regulators into how concerning these reporting limits are. While only 8.74% of analytes in the POTW samples were unreported at levels over 30ppt, a notably large 96.97% of the analytes sampled at PWS sites were unreported above 30ppt. This indicates an urgent need to reconsider and likely re-sample these sites for safety reasons. Further analysis should also consider the reporting limits as they pertain to specific analytes or chain-lengths, depending on health risks and upcoming regulation.

```
## [1] 8.737093
```

```
## [1] 96.96778
```

5 Summary and Conclusions

Summarize your major findings from your analyses in a few paragraphs. What conclusions do you draw from your findings? Relate your findings back to the original research questions and rationale.

The original purpose of this research was to synthesis the current publicly-available datasets on PFAS levels in North Carolina, begin building a bigger picture of the emerging contaminant PFAS presence, and identify areas of concern. The most notable findings from the cleaning process, exploration, and analysis are as follows:

1. The current data management structure for emerging contaminants in the state of North Carolina makes it prohibitively difficult for researchers and policymakers to grasp a comprehensive picture of PFAS occurance, distribution, and risk. Specifically, the static PDF report of POTW sample results, inconsistent names and formats for analytes, and confusing presentation of reporting limits (e.g., PWS dataset displaying limits in the same column as results) is a signifcant barrier to subsequent synthesis and analysis.

2. Despite signficant public and regulatory attention on the long-chain PFOA and PFOS analytes, this research clearly demonstrates that these are a small percentage of total PFAS present in North Carolina water. A narrow focus on these compounds causes an overlook of other water bodies that may pose significant health risks. Furthermore, it is clear that short-chain PFAS are the most abundant in these samples which should direct further research into associated health risks and inform remediation options because most treatment methods are ineffective at removing short-chain PFAS.

3. There is insufficient publicly-available data at this time to understand temporal and spatial trends, however the three-month sampling period for POTWs illustrated coinciding spikes and possibly an increasing trend over the summer months which needs further investigation.

4. The staggeringly high reporting limits (particularly for PWS samples) are a clear indication that any conclusions about risk, total PFAS, or appropriate remediation based on these datasets are likely misinformed. There is an urgent need for these sites to be re-tested in a manner that allows the lab to confidently report analyte levels at or above limits appropriate to their health risk (e.g, 10 ppt for long-chain PFAS).