WaterChemTS

Identifying factors that influence the sensitivity of water chemistry to climate variability

Bella Oleksy et al.

Last compiled on 07 December, 2021

Contents

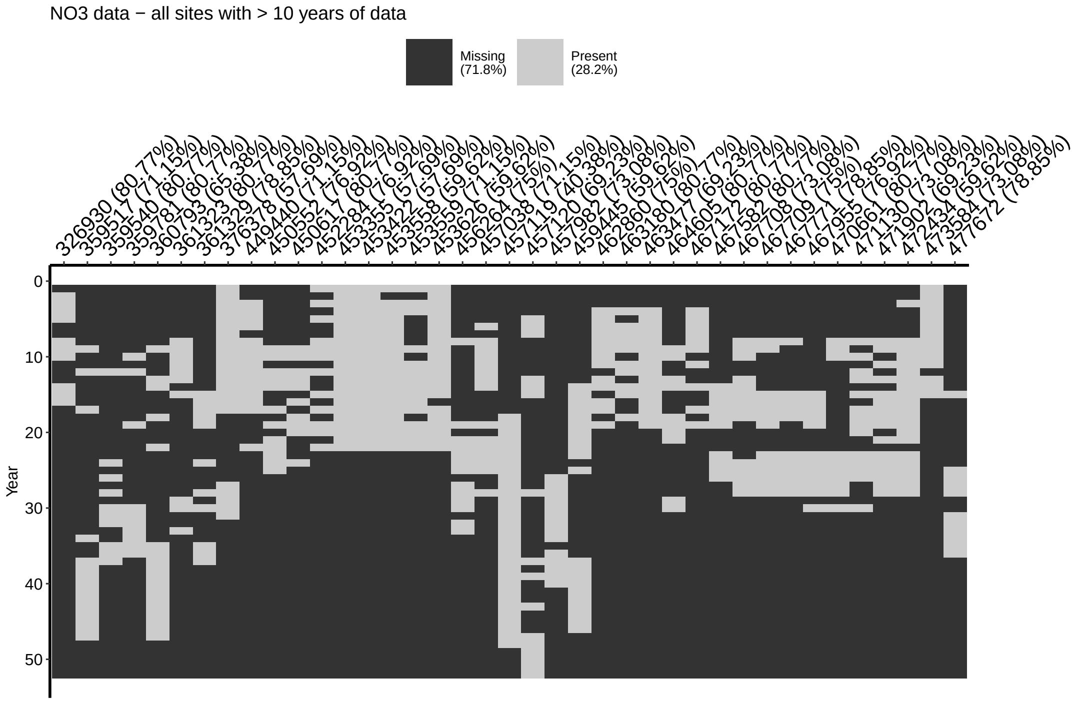
1 Initial data vis

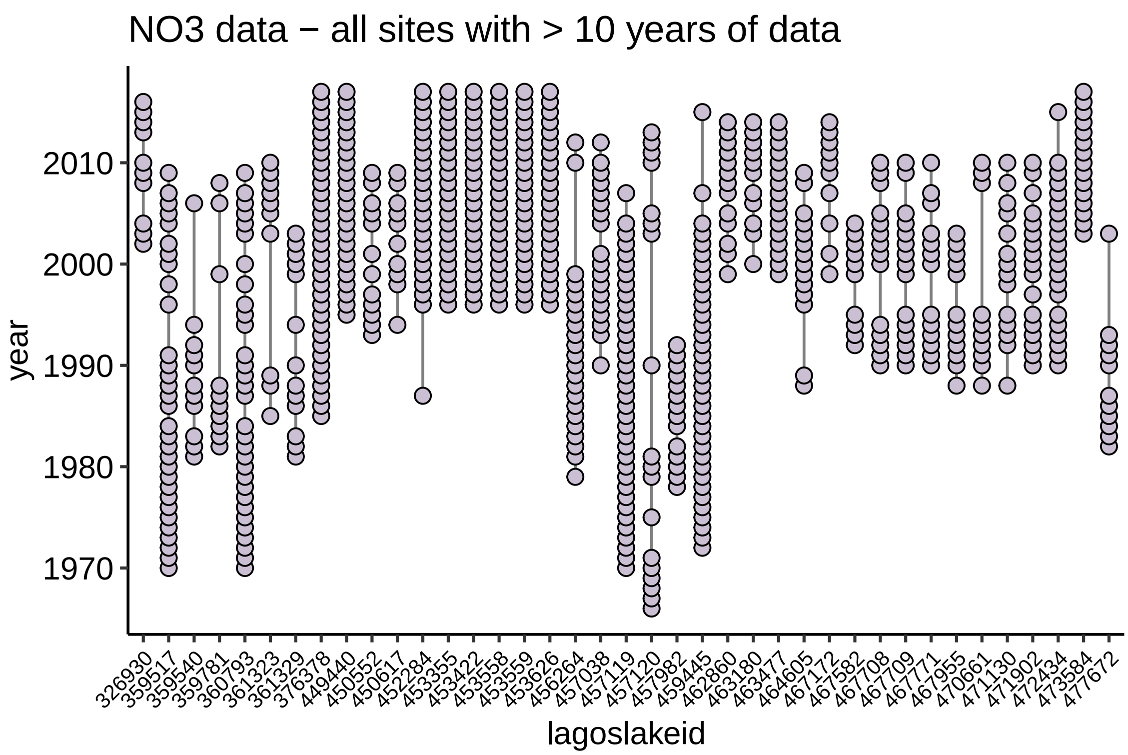
Before we dive into the analysis, we need to get a better sense of not just how many sites are there with > 10 years of data, but also the complete-ness of each time series. Ultimately this will determine what kind of timeseries models we use.

1.0.1 NO3

Currently in the database there are a total of 39 lakes with over 10 years of NO3 data. That seems great, but we need to take a closer look at how much missing data there is within each timeseries. Ultimately that will dictate our approach moving forward.

1.0.1.1 Visualize all site-years & missing data The next two plots are show a couple different ways that we can visualize the amount of missing data in the database. In the first one the y-axis is showing "year" based on row position, with more recent years at the top of the plot (zero) and older years near the bottom (50).





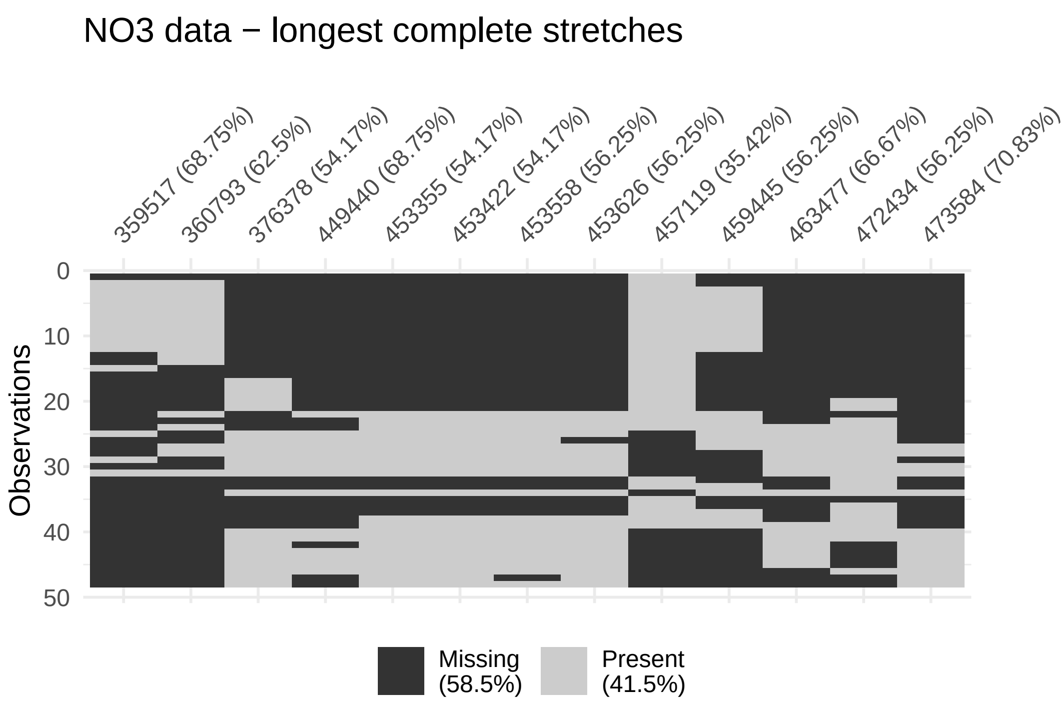
Basically, there are a lot of missing years, even for lakes with many years of data. Not the end of the world. What we want to know is how many lakes are there with the most complete, consecutive observations? I wrote a function for finding the longest complete stretch of annual observatiosn for each lake.

```
longestCompleteStretch <- function(x) {
  with(rle(!is.na(x)), max(lengths[values]))
}
```

```
#Calculate the longest stretch of complete observations for each lake
NO3_complete<-NO3_LT_wide %>%
  summarise_at(vars(1:ncol(.)), longestCompleteStretch) %>%
  pivot_longer(1:ncol(.)) %>%
  rename(longestCompleteStretch=value,
         lagoslakeid=name) %>%
  filter(longestCompleteStretch>=9)

#Pull out the names of lakes where 100% of the annual timeseries is complete
NO3_complete_10year_names <- NO3_complete %>%
  filter(longestCompleteStretch>=10) %>%
  pivot_wider(names_from=lagoslakeid, values_from=longestCompleteStretch) %>%
  names()
```

1.0.1.2 Visualize all longest complete stretches of data Now we are left with 13 lakes. They still vary in the timeperiod of observation, with a few of these sites starting observations in the 1970s, but possibly phasing out monitoring in more recent years. Alternatively, the database hasn't been updated yet.



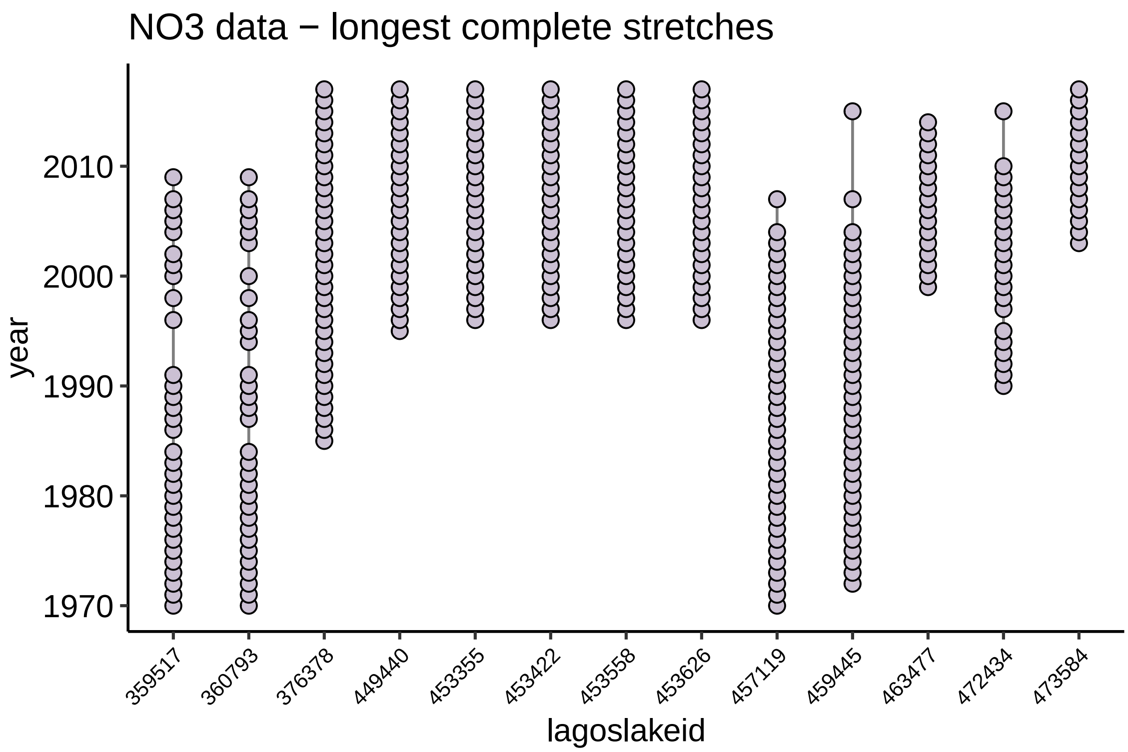


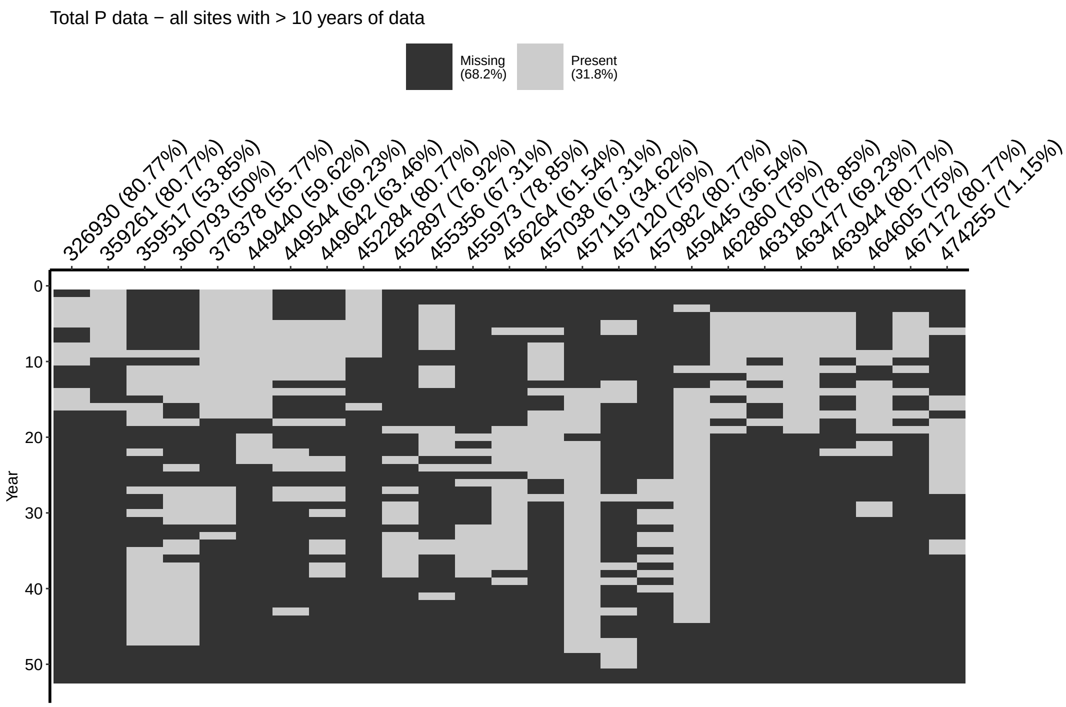
Table 1: Lakes with with >= 10 years of NO3 data, and no year gaps for a minimum of 10 years

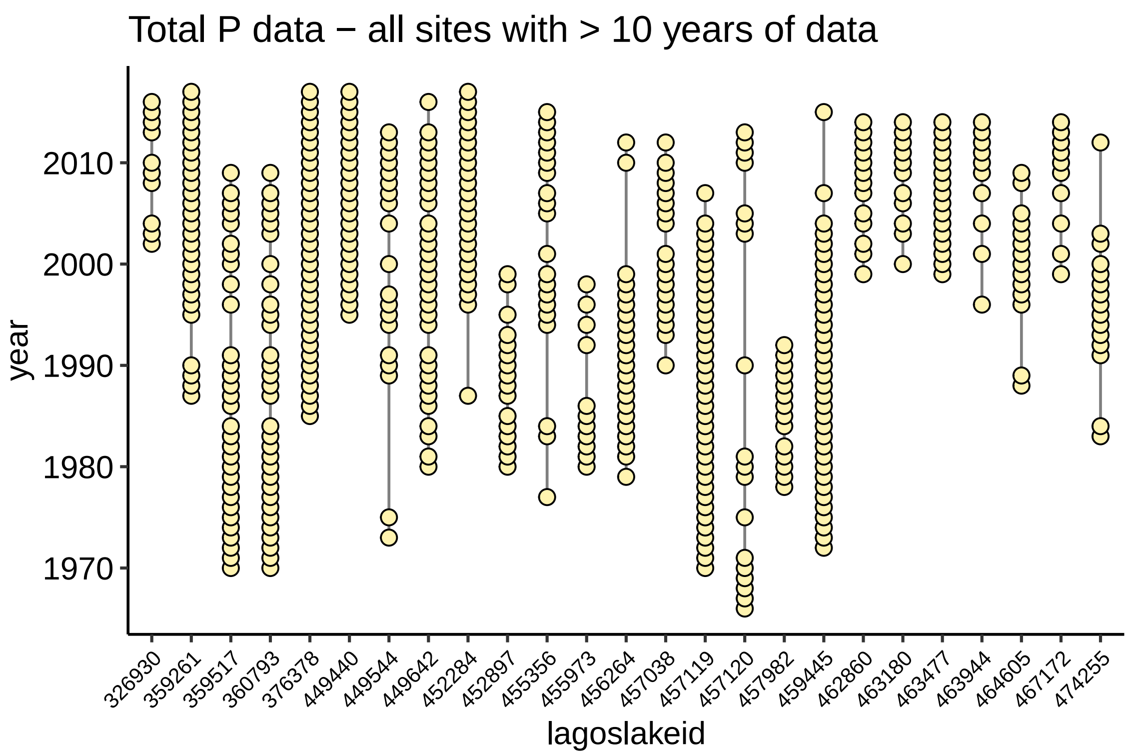
lagoslakeid	Lake name	Lat	Long	Years of data (n)	First year	Last year
359517	Carter Lake Reservoir	40.33524	-105.2179	31	1970	2009
360793	Horsetooth Reservoir	40.55828	-105.1595	31	1970	2009
376378	Pueblo Reservoir	38.26696	-104.7504	33	1985	2017
449440	Wolford Mountain Reservoir	40.14429	-106.4040	23	1995	2017
453355	NA	37.62763	-107.5831	22	1996	2017
453422	NA	37.70820	-107.5525	22	1996	2017
453558	NA	37.62047	-107.5861	22	1996	2017
453626	Eldorado Lake	37.71180	-107.5436	22	1996	2017
457119	Topaz Lake	38.67339	-119.5389	36	1970	2007
459445	Lake Koocanusa	48.91013	-115.1948	35	1972	2015
463477	Lake Walcott	42.62895	-113.2179	16	1999	2014
472434	Mowich Lake	46.93827	-121.8621	21	1990	2015
473584	Upper Klamath Lake; Melvins Pond	42.35042	-121.8624	15	2003	2017

1.0.2 TP

Currently in the database there are a total of 25 lakes with over 10 years of TP data. That seems great, but we need to take a closer look at how much missing data there is within each timeseries. Ultimately that will dictate our approach moving forward.

1.0.2.1 Visualize all site-years & missing data The next two plots are show a couple different ways that we can visualize the amount of missing data in the database. In the first one the y-axis is showing "year" based on row position, with more recent years at the top of the plot (zero) and older years near the bottom (50).





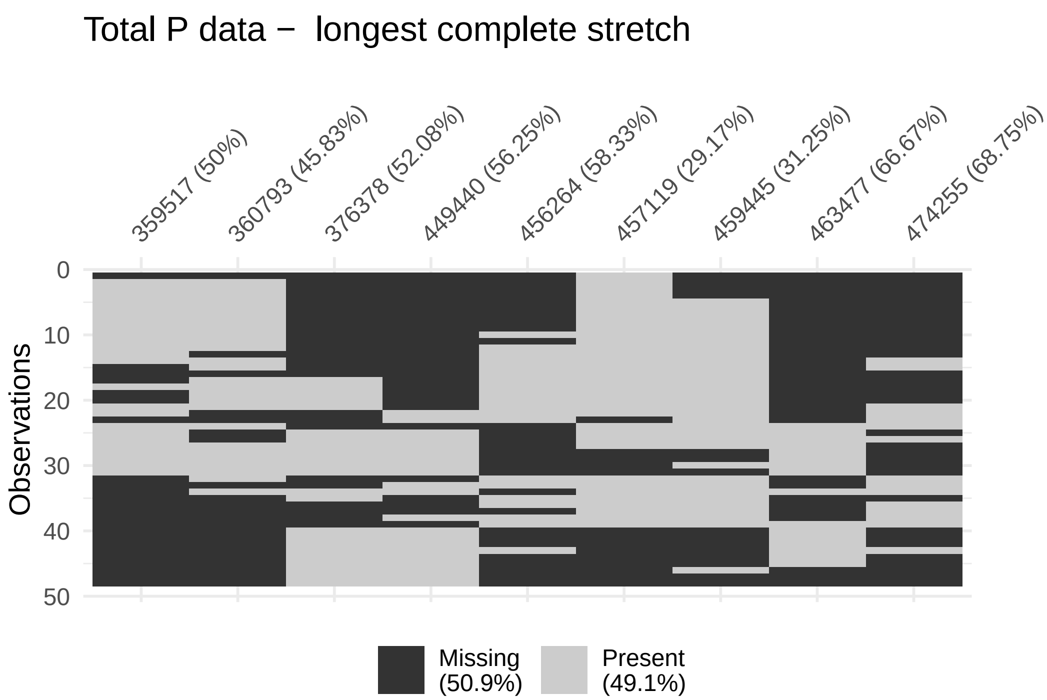
Basically, there are a lot of missing years, even for lakes with many years of data. Not the end of the world. What we want to know is how many lakes are there with the most complete, consecutive observations?

```
#Calculate the longest stretch of complete observations for each lake
TP_complete<-TP_LT_wide %>%
  summarise_at(vars(1:ncol(.)), longestCompleteStretch) %>%
  pivot_longer(1:ncol(.)) %>%
  rename(longestCompleteStretch=value,
```

```
lagoslakeid=name) %>%
  filter(longestCompleteStretch>=9)

#Pull out the names of lakes where 100% of the annual timeseries is complete
TP_complete_10year_names <- TP_complete %>%
  filter(longestCompleteStretch>=10) %>%
  pivot_wider(names_from=lagoslakeid, values_from=longestCompleteStretch) %>%
  names()
```

1.0.2.2 Visualize all longest complete stretches of data Now we are left with 9 lakes. They still vary in the timeperiod of observation, with a few of these sites starting observations in the 1970s, but possibly phasing out monitoring in more recent years. Alternatively, the database hasn't been updated yet.



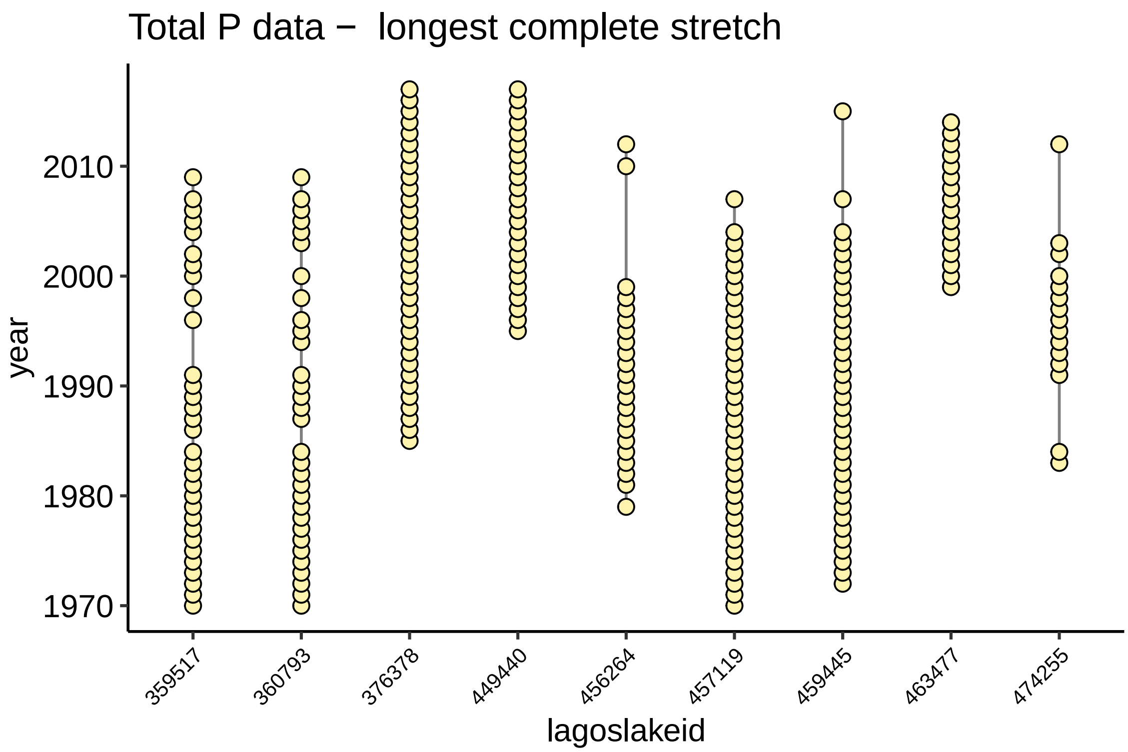


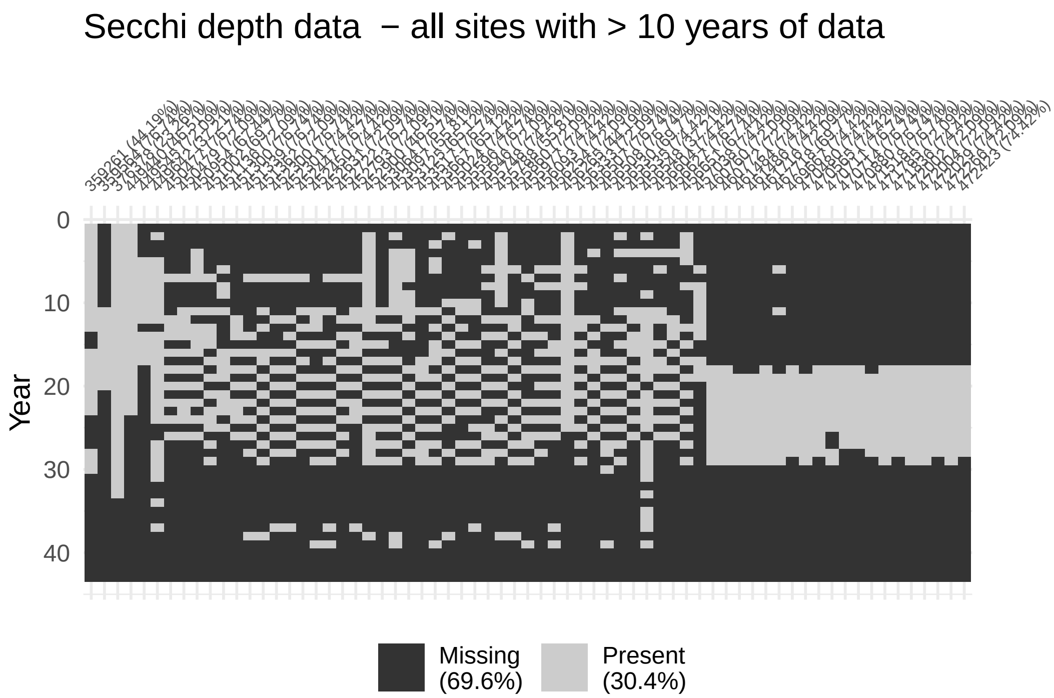
Table 2: Lakes with with >= 10 years of TP data, and no year gaps for a minimum of 10 years

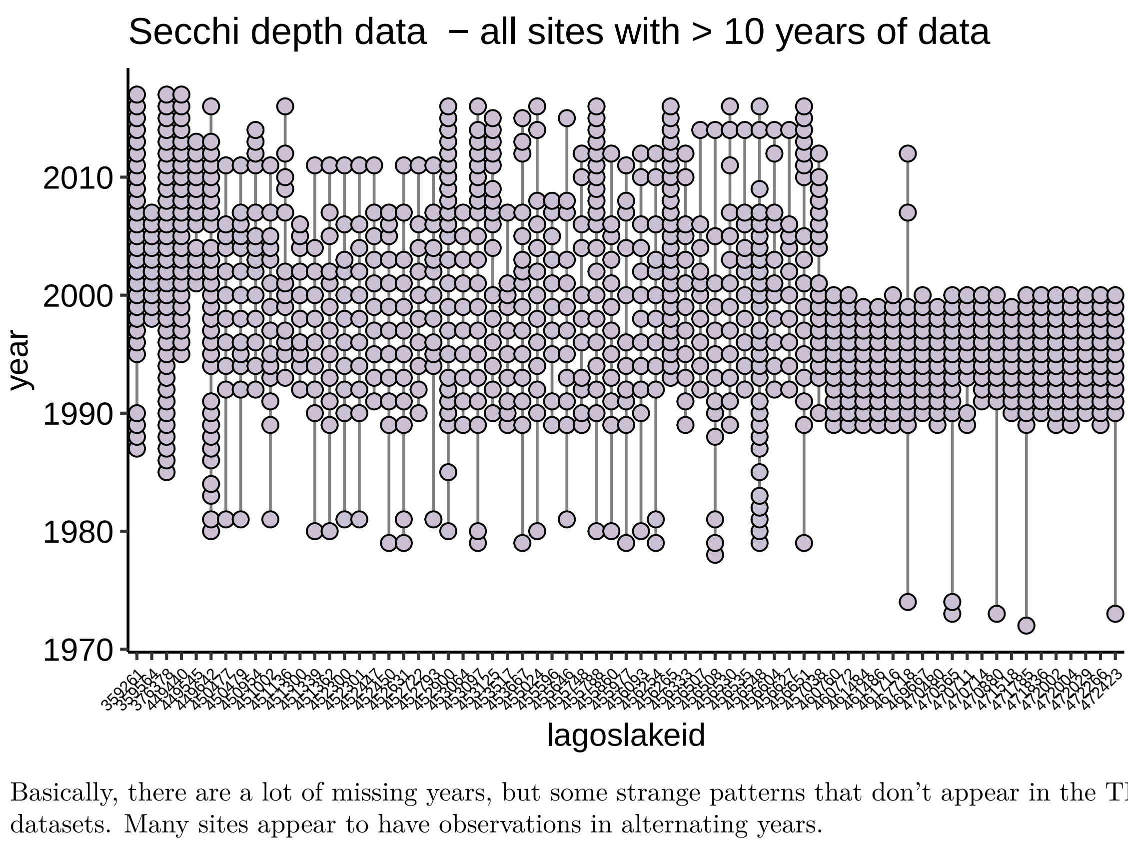
lagoslakeid	Lake name	Lat	Long	Years of data (n)	First year	Last year
359517	Carter Lake Reservoir	40.33524	-105.2179	31	1970	2009
360793	Horsetooth Reservoir	40.55828	-105.1595	31	1970	2009
376378	Pueblo Reservoir	38.26696	-104.7504	33	1985	2017
449440	Wolford Mountain Reservoir	40.14429	-106.4040	23	1995	2017
456264	Deer Creek Reservoir	40.44337	-111.4934	22	1979	2012
457119	Topaz Lake	38.67339	-119.5389	36	1970	2007
459445	Lake Koocanusa	48.91013	-115.1948	35	1972	2015
463477	Lake Walcott	42.62895	-113.2179	16	1999	2014
474255	Crater Lake	42.94195	-122.1073	15	1983	2012

1.0.3 Secchi depth

Currently in the database there are a total of 67 lakes with over 10 years of secchi depth data. That seems great, but we need to take a closer look at how much missing data there is within each timeseries. Ultimately that will dictate our approach moving forward.

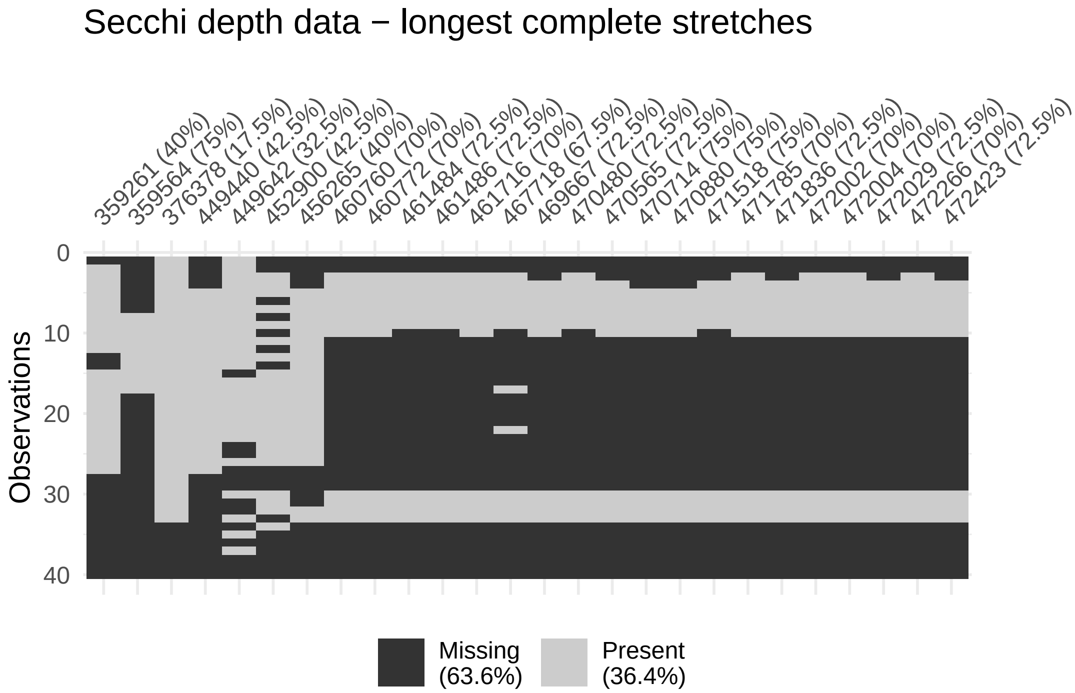
1.0.3.1 Visualize all site-years & missing data The next two plots are show a couple different ways that we can visualize the amount of missing data in the database. In the first one the y-axis is showing "year" based on row position, with more recent years at the top of the plot (zero) and older years near the bottom (50).

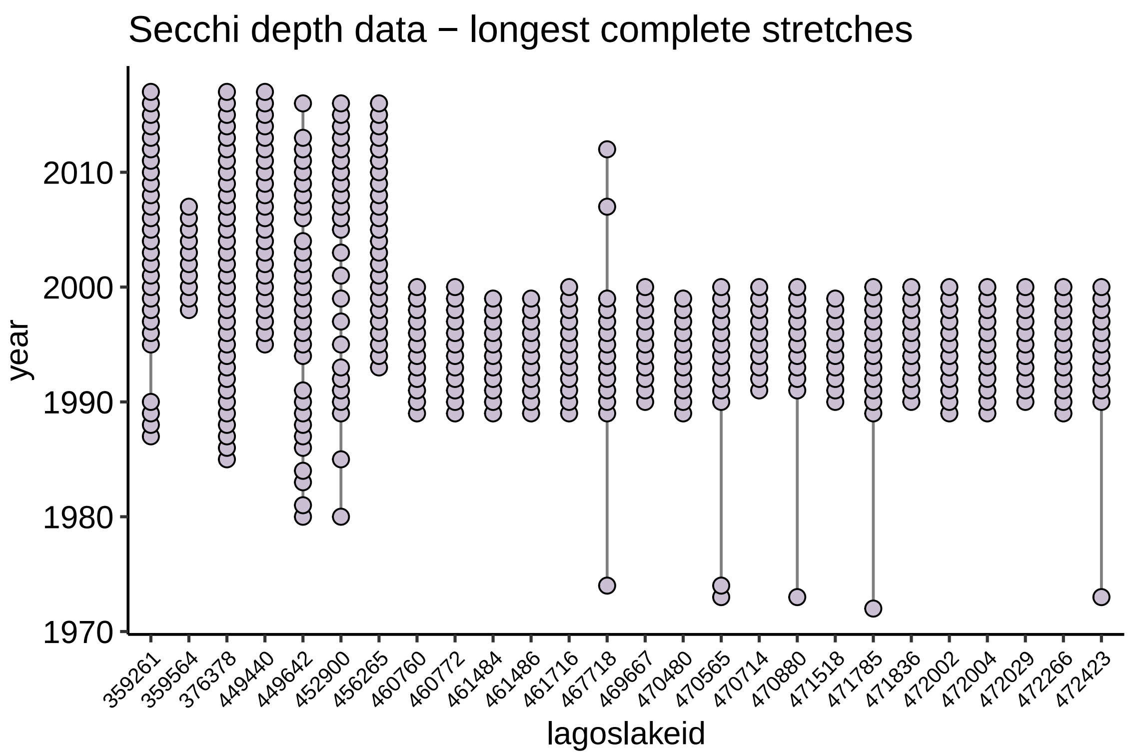




Basically, there are a lot of missing years, but some strange patterns that don't appear in the TP and NO3 datasets. Many sites appear to have observations in alternating years.

1.0.3.2 Visualize all longest complete stretches of data Now we are left with 26 lakes.





We are left with a much higher sample size but interestingly, for a bunch of these lakes the monitoring stopped in 2000. What gives?

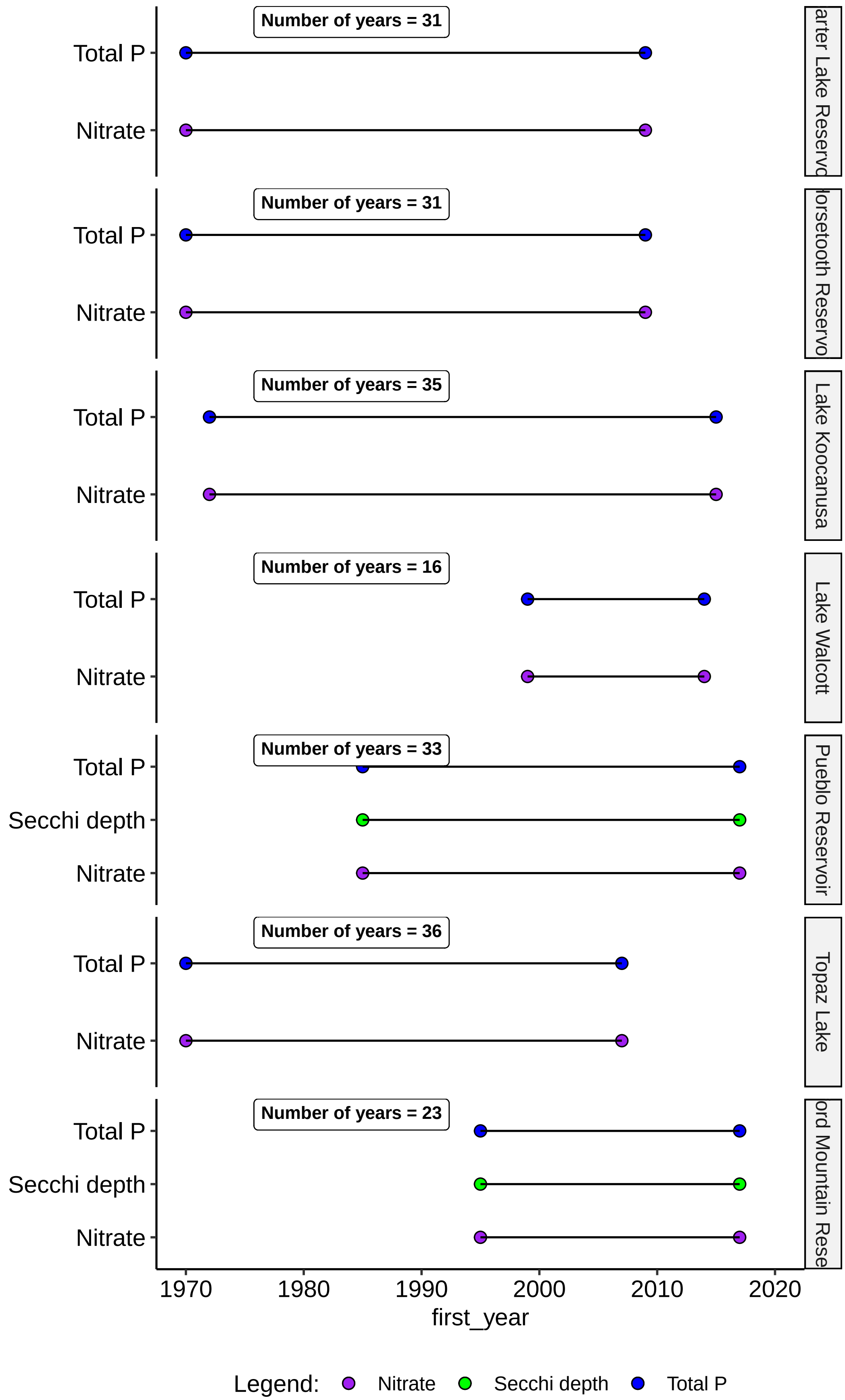
Table 3: Lakes with with >= 10 years of SECCHI data, and no year gaps for a minimum of 10 years

lagoslakeid	Lake name	Lat	Long	Years of data (n)	First year	Last year
359261	Summit Lake	40.54543	-106.6824	27	1987	2017
359564	Lake Estes	40.37559	-105.4930	10	1998	2007
376378	Pueblo Reservoir	38.26696	-104.7504	33	1985	2017
449440	Wolford Mountain Reservoir	40.14429	-106.4040	23	1995	2017
449642	Lake Granby; Arapaho Bay	40.16145	-105.8645	30	1980	2016
452900	Red Fleet Reservoir	40.57942	-109.4306	24	1980	2016
456265	Jordanelle Reservoir	40.61890	-111.4086	24	1993	2016
460760	Black Lake	48.56195	-117.6260	12	1989	2000
460772	Osoyoos Lake	48.97751	-119.4371	12	1989	2000
461484	Big Meadow Lake	48.72777	-117.5577	11	1989	1999
461486	Lake Thomas	48.62428	-117.5403	11	1989	1999
461716	Lake Wenatchee	47.82266	-120.7772	12	1989	2000
467718	Lacamas Lake	45.61647	-122.4252	14	1974	2012
469667	Nahwatzel Lake	47.24239	-123.3332	11	1990	2000
470480	Lake Wooten	47.46730	-122.9816	11	1989	1999
470565	Lake Martha	48.16720	-122.3383	13	1973	2000
470714	Lake Roesiger	47.99461	-121.9113	10	1991	2000
470880	Lake Alice	47.53178	-121.8904	11	1973	2000
471518	Ward Lake	47.00877	-122.8754	10	1990	1999
471785	Phillips Lake	47.25028	-122.9595	13	1972	2000
471836	Spanaway Lake	47.11013	-122.4481	11	1990	2000
472002	Mason Lake	47.34054	-122.9563	12	1989	2000
472004	Lake Samish	48.66367	-122.3863	12	1989	2000
472029	Lake Limerick	47.28613	-123.0453	11	1990	2000
472266	Lake Saint Clair	47.00029	-122.7200	12	1989	2000
472423	Bosworth Lake	48.04336	-121.9707	12	1973	2000

2 Common lakes

Lastly, this table shows all the lakes with long-term data of any kind (TP, NO3, Secchi depth) as a way of seeing if/which lakes have multiple parameters. For instance, we talked about looking at changing NO3:TP ratios.

For cleaner plotting, I am only showing the lakes where we have both NO3 & TP data or all three.



Number of years = 31
Total P
Nitrate
arter Lake Reservo
Number of years = 31
Total P
Nitrate
Horsetooth Reservo
Number of years = 35
Total P
Nitrate
Lake Koocanusa
Number of years = 16
Total P
Nitrate
Lake Walcott
Number of years = 33
Total P
Secchi depth
Nitrate
Pueblo Reservoir
Number of years = 36
Total P
Nitrate
Topaz Lake
Number of years = 23
Total P
Secchi depth
Nitrate
ord Mountain Rese
1970
1980
1990
2000
2010
2020
first_year
Legend:
Nitrate
Secchi depth
Total P