Reply to questions in the last note

Mengran Li

2021/3/2

Do we have data for almost every day?

There are $365 \times 40 + 10 = 14610$ days, and therefore there should be $14610 \times 671 = 9803310$ in the case of no NA. Actually, we have 9559629 observations. I have checked and confirmed that these was no NA. The proportion is $9559629/9803310 = 97.51\%$, which is acceptable in my view.

Are there fewer locations or rivers/ stream in the platau area? & Common geographical features?

I have written a web map https://gla-my.sharepoint.com/:u:/g/personal/2592713l_student_gla_ac_uk/Ee vYgJRRfHVOjcxnTMft8UwBplq8Ch1tAJXk4W-THV6tug?e=2dfM0s.

My conclusion is, rivers in the platau area is really fewer, which may also cause the shortage of observations here. As what I saw in the UCGS with full observations, there is same distribution in this area.

I think another reasonable interpretation that we see fewer dots in the area in higher quantile maps is, the Q values here are smaller than other area so that when we select threshold, almost none is left (while smaller quantiles opposite).

My question is, whether we need to add some observations here. If we do, however, how can we choose sites added and how to explain the reason.

Given by the geographical features, I have some thoughts.

- Bivariate extreme model
- Another possible division about region (Streamflow feature, Climate, etc.), I mean, more close to graph we draw.
- Maybe establish a function of distance (not sure)

Season

My former thought is to divide the months into two season based on graph, one of which is stable and another fluctuates periodically, so that we can explore their period effect or whether they have some relationship further. But now I gave up the idea. US has almost all types of climate, thus there are complex precipitation situation in different regions.

However, I notice that their changing in precipitation usually can be described by unit of season because of monsoon and atmospheric pressure. Precipitation would bring important influence on streamflow. Therefore, I think we may can explore the trend of seasons.

I divide the months into seasons by The official travel site of the USA https://www.visittheusa.co.uk/info/w eather-seasons
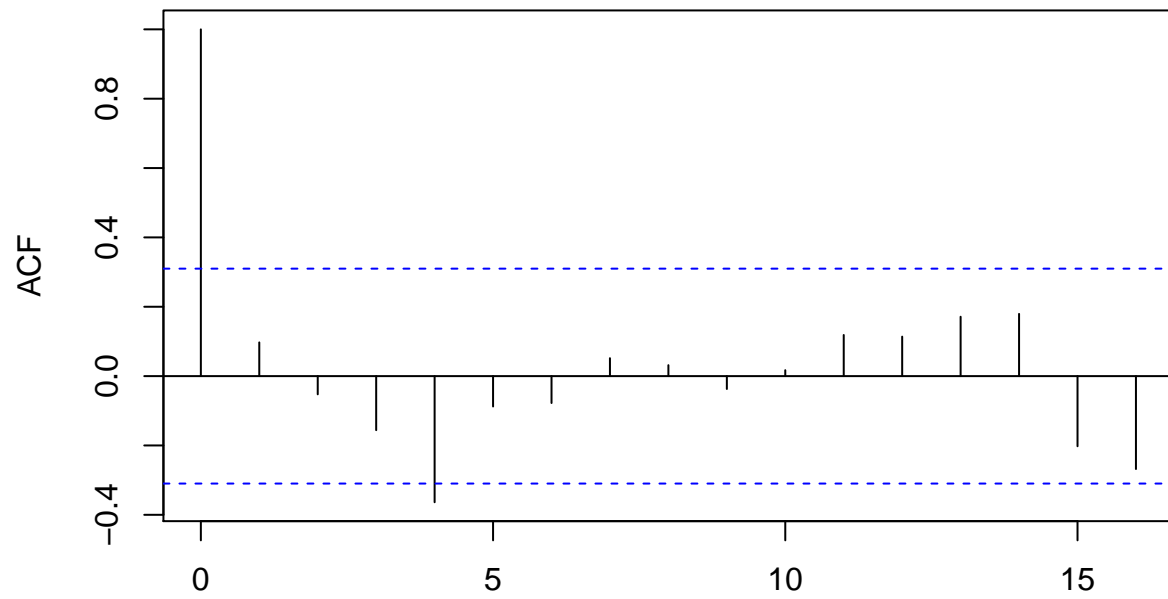
- Summer: June - August
- Fall: September - November

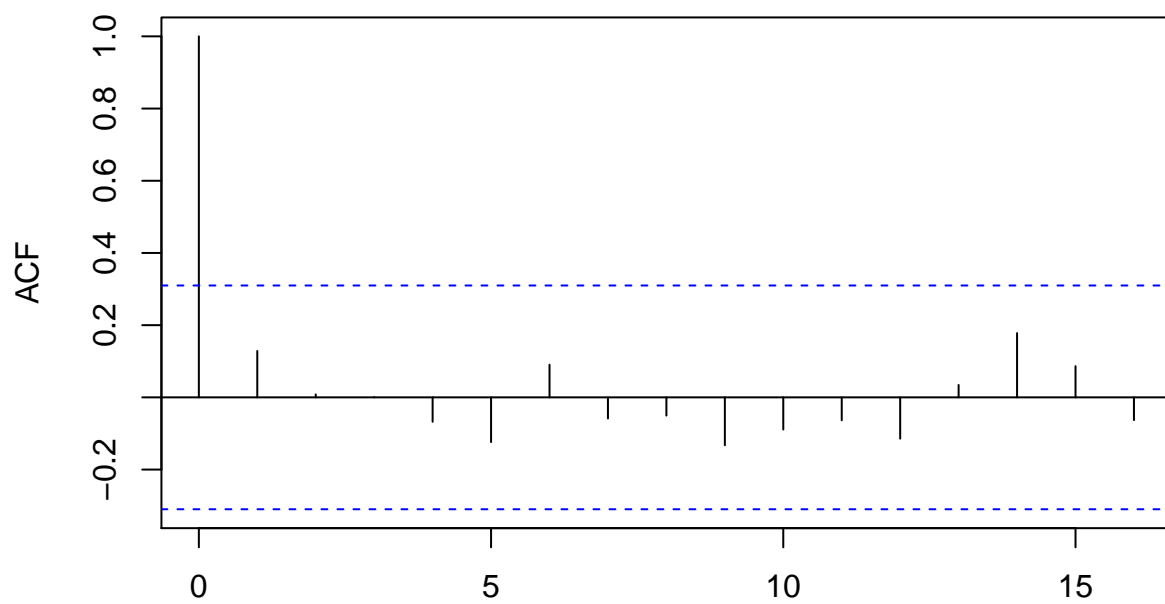- Winter: December - February

- Spring: March - May
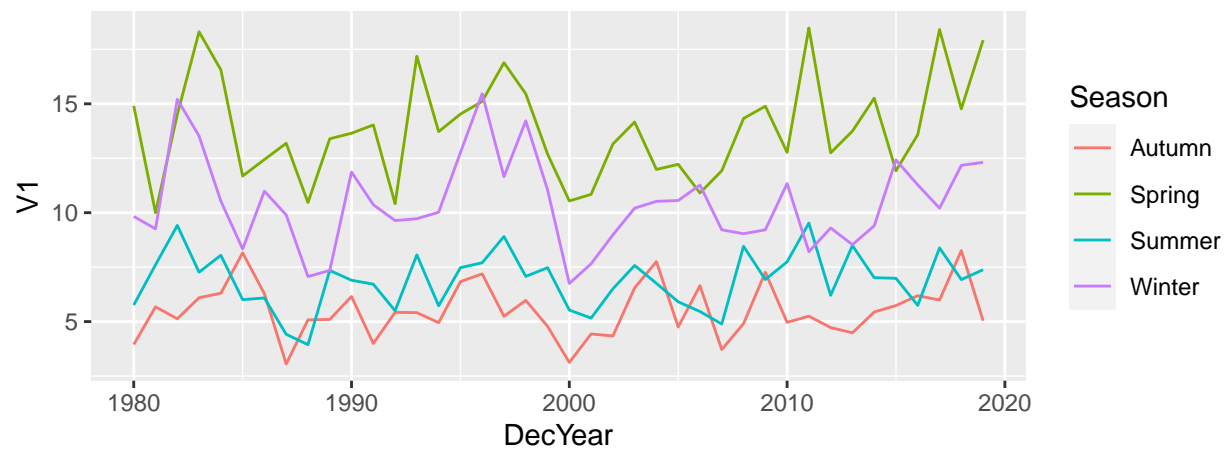
The plot is not smoothed. I think it is better than 12-months plot. They are not so messy than former plot. I try to find their trend.
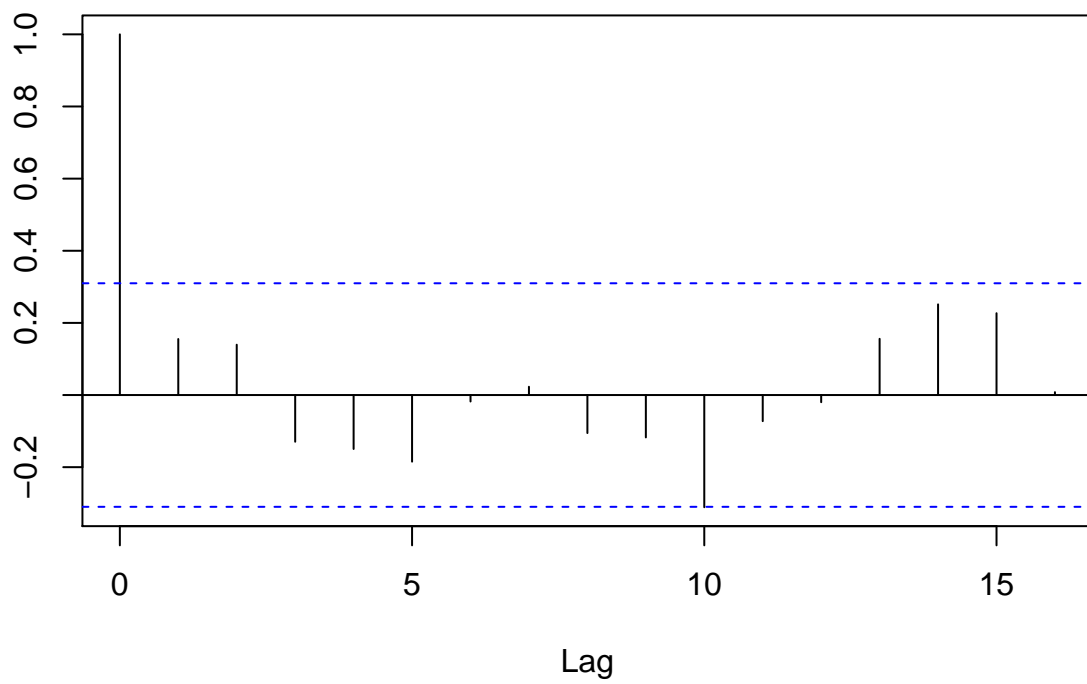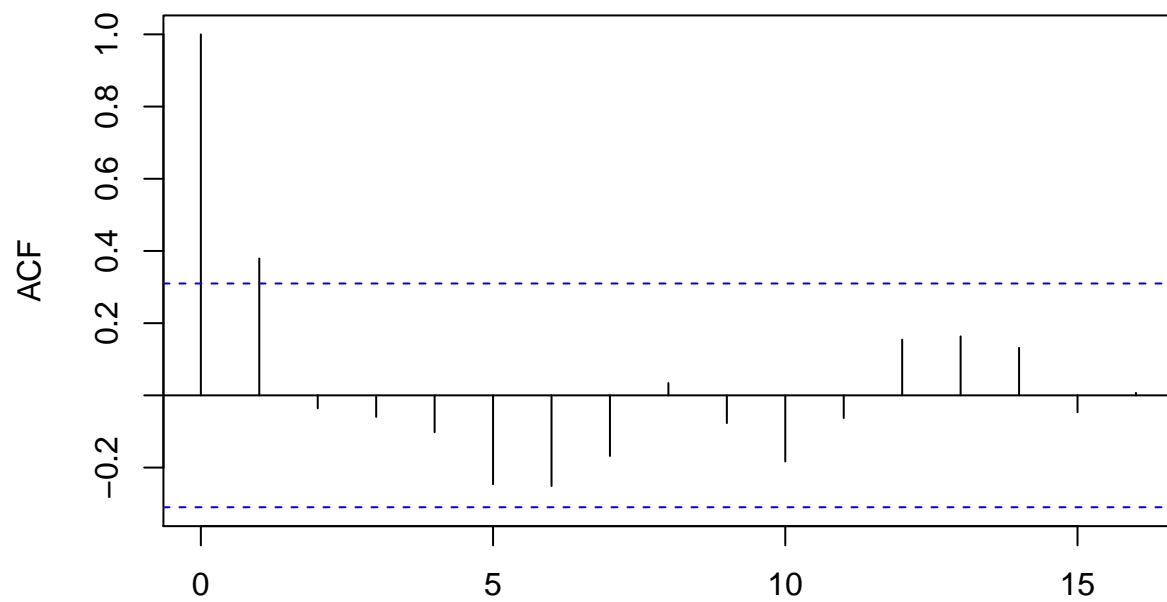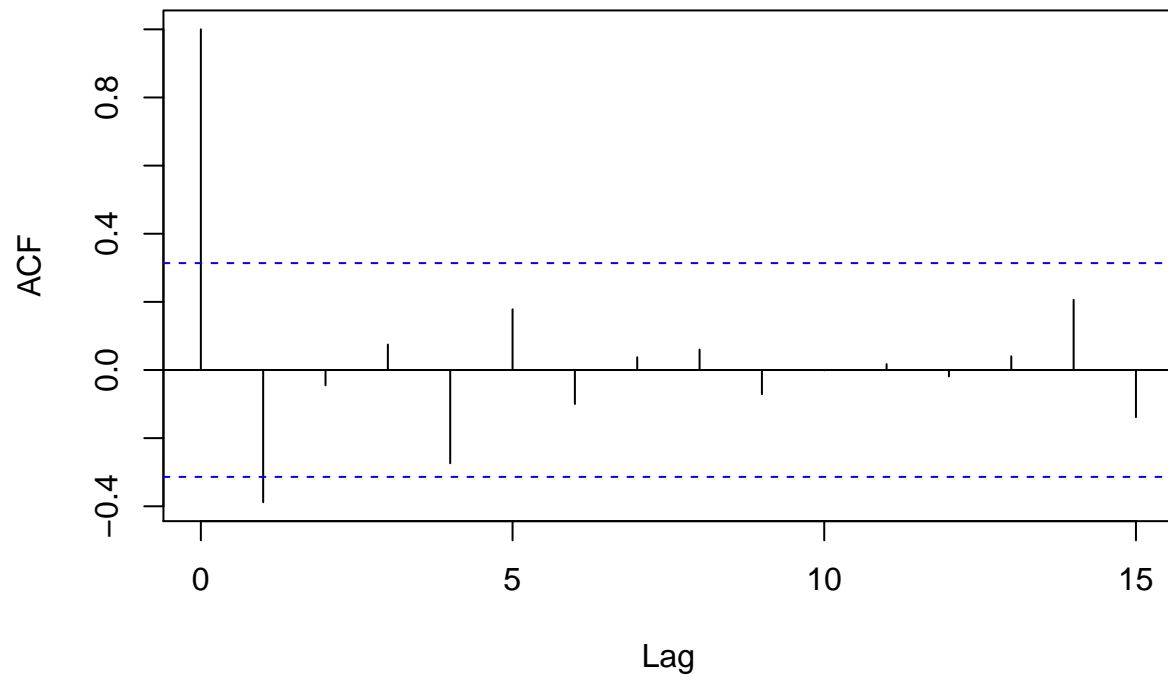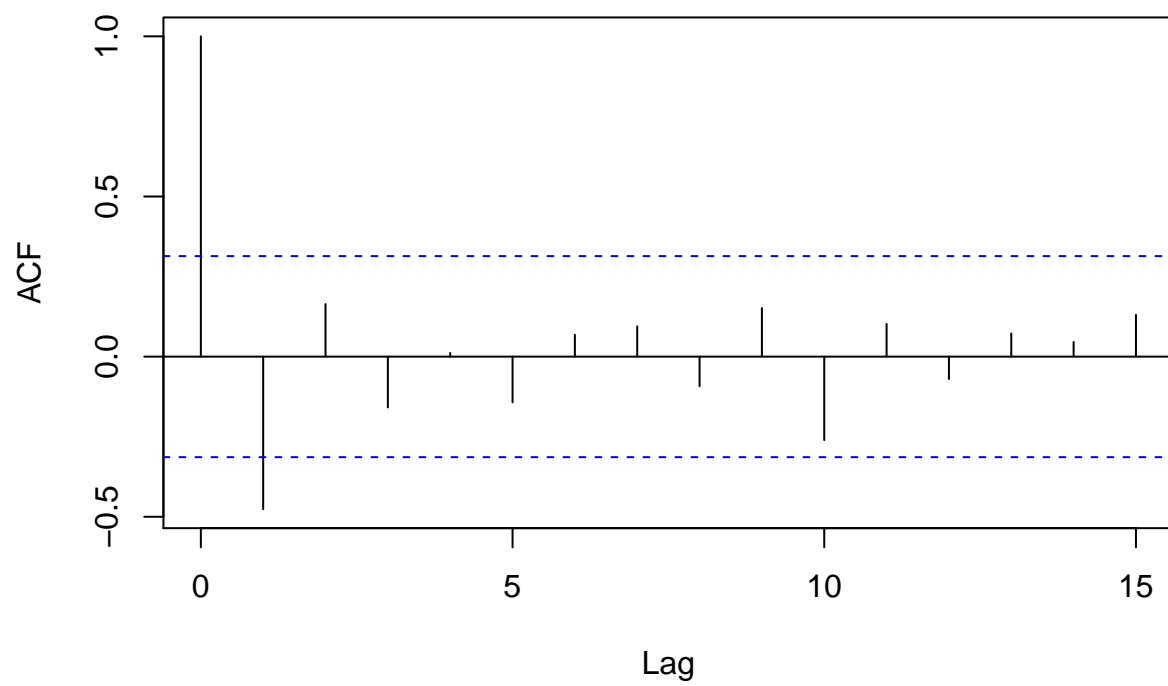
Spring

Summer

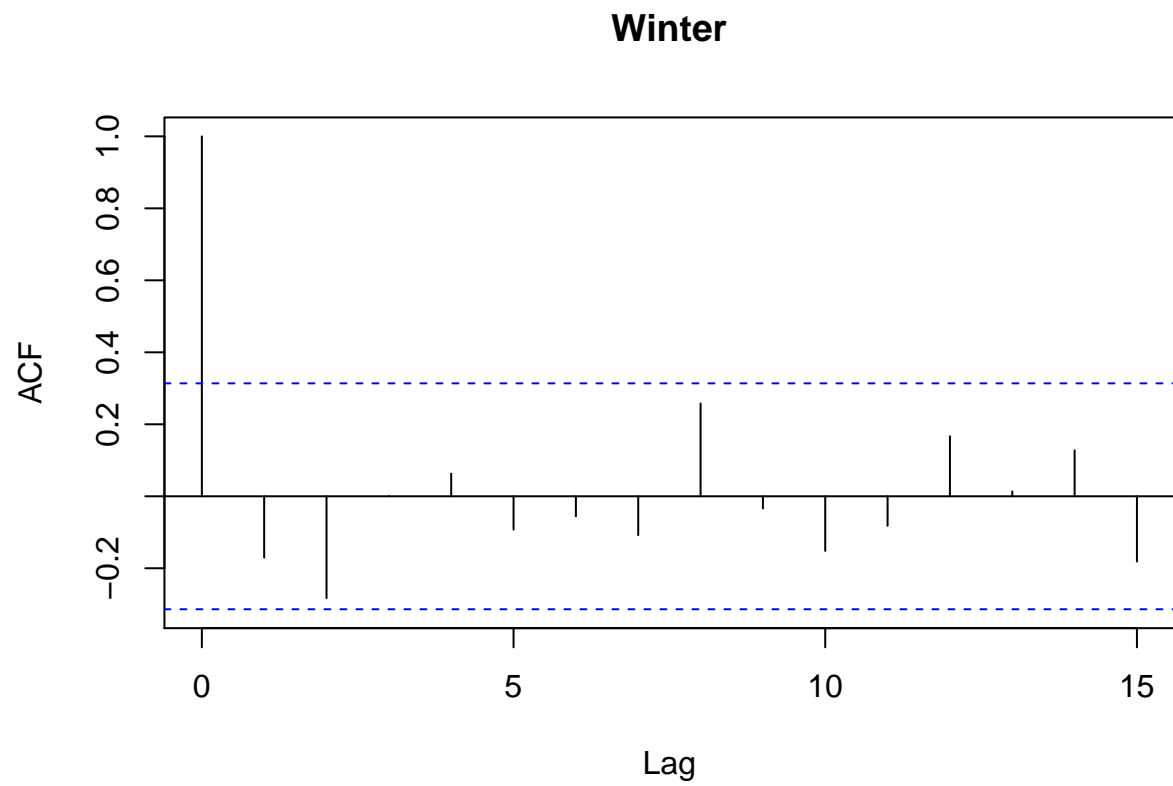**Autumn**

Lag

Winter

Summer is stable. Take the difference to Spring, Autumn and Winter.
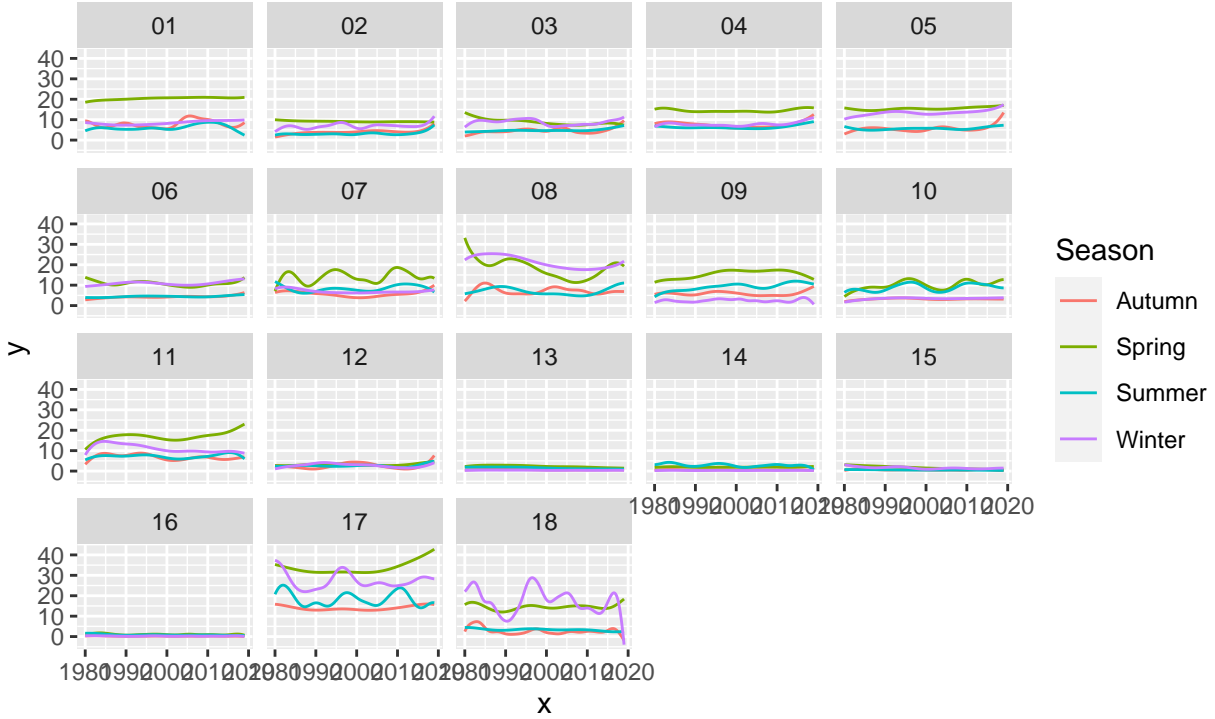
Spring

Lag

ACF

Autumn

Winter

Spring and Autumn are still 1 lag autocorrelation.

Winter is stable.

Show quantiles by location on maps.

I calculated statistics of each region. And the table here is a summary of each statistics. The max value is the reason that the colours seem very close in the map.

```
##    stat        Mean         q0.05          q0.1          q0.2          q0.5
## 1:  max 243.5953152 3.454655e+01 4.162576e+01 5.493468e+01 1.059050e+02
## 2: Mean   9.0692532 1.182500e+00 1.433792e+00 1.940148e+00 4.461139e+00
## 3:  min   0.0157077 2.903672e-05 2.903672e-05 2.903672e-05 5.936966e-05
## 4:   Q1   1.1022595 2.852429e-02 4.445745e-02 7.743566e-02 2.834469e-01
## 5:   Q2   3.6039332 1.953862e-01 2.632617e-01 4.252478e-01 1.444159e+00
## 6:   Q3   9.3236730 8.495054e-01 1.082507e+00 1.575833e+00 4.282614e+00
##            q0.7         q0.9        q0.95          Sd
## 1: 2.487918e+02 6.569508e+02 8.936797e+02 296.81011479
## 2: 8.210434e+00 2.031211e+01 3.105622e+01  15.71117683
## 3: 2.246666e-04 2.246666e-04 2.246666e-04   0.02312148
## 4: 6.518970e-01 2.203798e+00 3.950200e+00   2.60060384
## 5: 2.803368e+00 7.192479e+00 1.202545e+01   7.57062264
## 6: 7.603073e+00 1.950748e+01 3.284754e+01  18.37518899
```
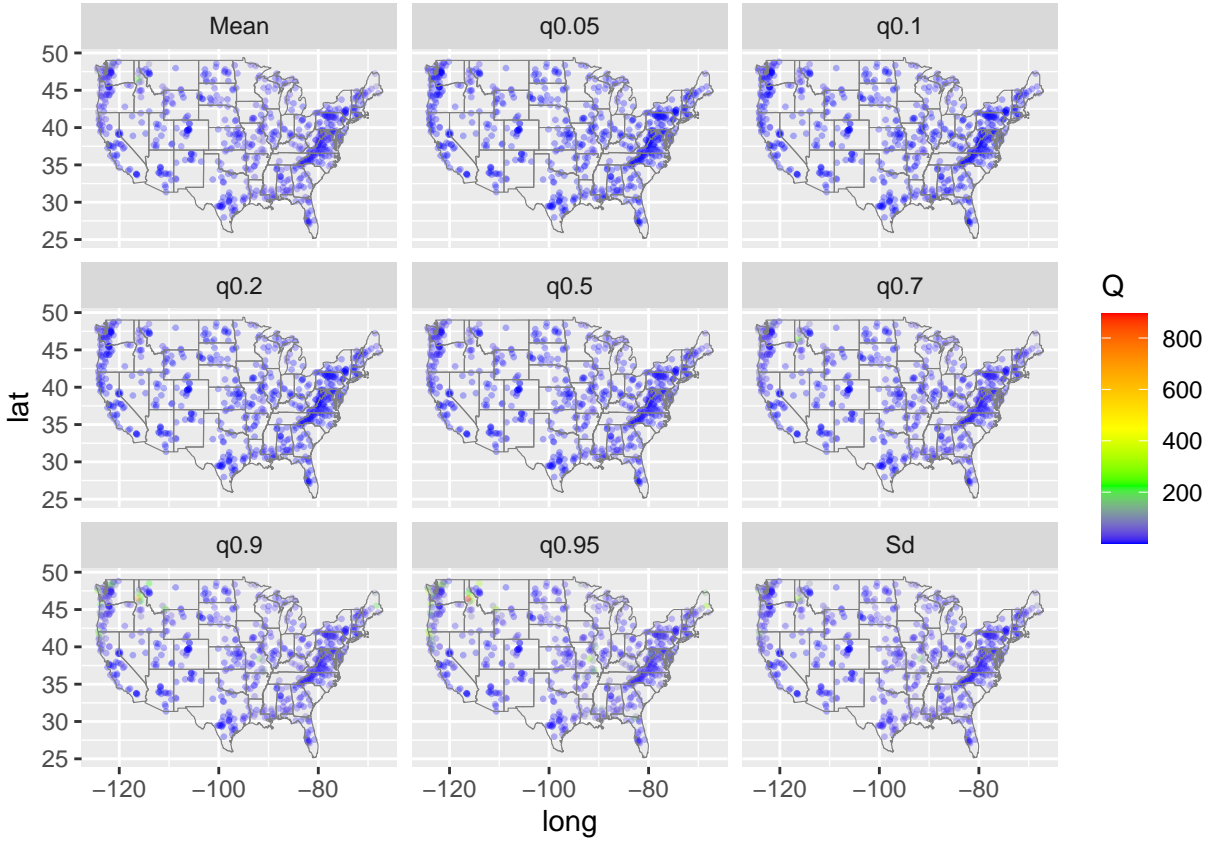
All plots are available in the github https://github.com/MengranLi-git/StreamFlow/tree/main/Plot. They are sorted well so that easy to find. And they look better than Rmarkdown. Plots are displayed differently from the window in R.