Michelle Visscher DATA423 A4

Code

Question 2: Control Charts

Step 1: Loading Files.

The file monitor.csv contains comma separated data. The columns are Timestamp - the time-stamp of a model prediction being run ProcessMemory - the allocated memory (MB) of the relevant server process Prediction - the value predicted by the model PredictionTimeMS - the duration of the prediction task in milliseconds

Hide

```
monitor_data <- read.csv("monitor.csv")
skim(monitor_data)
```

```
── Data Summary ────────────────────────
                           Values
Name                       monitor_data
Number of rows             16069
Number of columns          4
_____________________
Column type frequency:
  character                1
  numeric                  3
_____________________
Group variables            None
```

skim_variable	n_missing	complete_rate	…	…	em…	n_unique	whitespace
<chr>	<int>	<dbl>	<int>	<int>	<int>	<int>	<int>
1 Timestamp	0	1	19	19	0	16056	0

1 row

skim_variable	n_missing	complete_rate	mean	sd	p0	
<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	
1 ProcessMemory	0	1	9.071349	0.4034112	7.326019	8.7
2 Prediction	0	1	52.208205	9.4174825	36.010109	44.1
3 PredictionTimeMS	0	1	360.063187	81.6698211	24.343181	305.2

3 rows | 1-9 of 11 columns

Add a day-of-the-year column or something similar to marks the days.

Hide

```
monitor_data$day_of_year <- yday(ymd_hms(monitor_data$Timestamp))
head(monitor_data)
```

	Timestamp <chr>	ProcessMemory <dbl>	Prediction <dbl>	PredictionTimeMS <dbl>	day_of_year <dbl>
1	2021-02-01 13:00:24	8.785348	51.19133	476.4944	32
2	2021-02-01 13:10:58	9.549011	44.38603	315.2675	32
3	2021-02-01 13:26:37	8.767273	59.62085	310.9753	32
4	2021-02-01 13:29:13	9.661670	63.14264	348.2056	32
5	2021-02-01 13:35:20	9.523878	40.35872	494.6693	32
6	2021-02-01 13:59:20	9.303197	53.44916	393.1038	32

6 rows

Hide

```
tail(monitor_data)
```

	Timestamp <chr>	ProcessMem... <dbl>	Prediction <dbl>	PredictionTimeMS <dbl>	day_of_year <dbl>
16064	2021-06-01 11:18:35	9.315040	41.30944	429.4449	152
16065	2021-06-01 11:32:00	8.714421	44.98940	361.0256	152
16066	2021-06-01 11:37:04	9.278243	36.52788	343.2776	152
16067	2021-06-01 11:40:57	8.756996	43.42669	305.5567	152
16068	2021-06-01 11:45:06	9.094910	49.13386	454.5351	152
16069	2021-06-01 11:45:52	8.713085	36.69393	305.9007	152

6 rows

Hide

```
NA
```

Day of year starts at 32 because the first day is 1st Feburary 2021 - meaning the 32nd day of the year The last day of the data set is 1st June 2021, meaning the 152nd day of the year. In total there are 16069 observations.

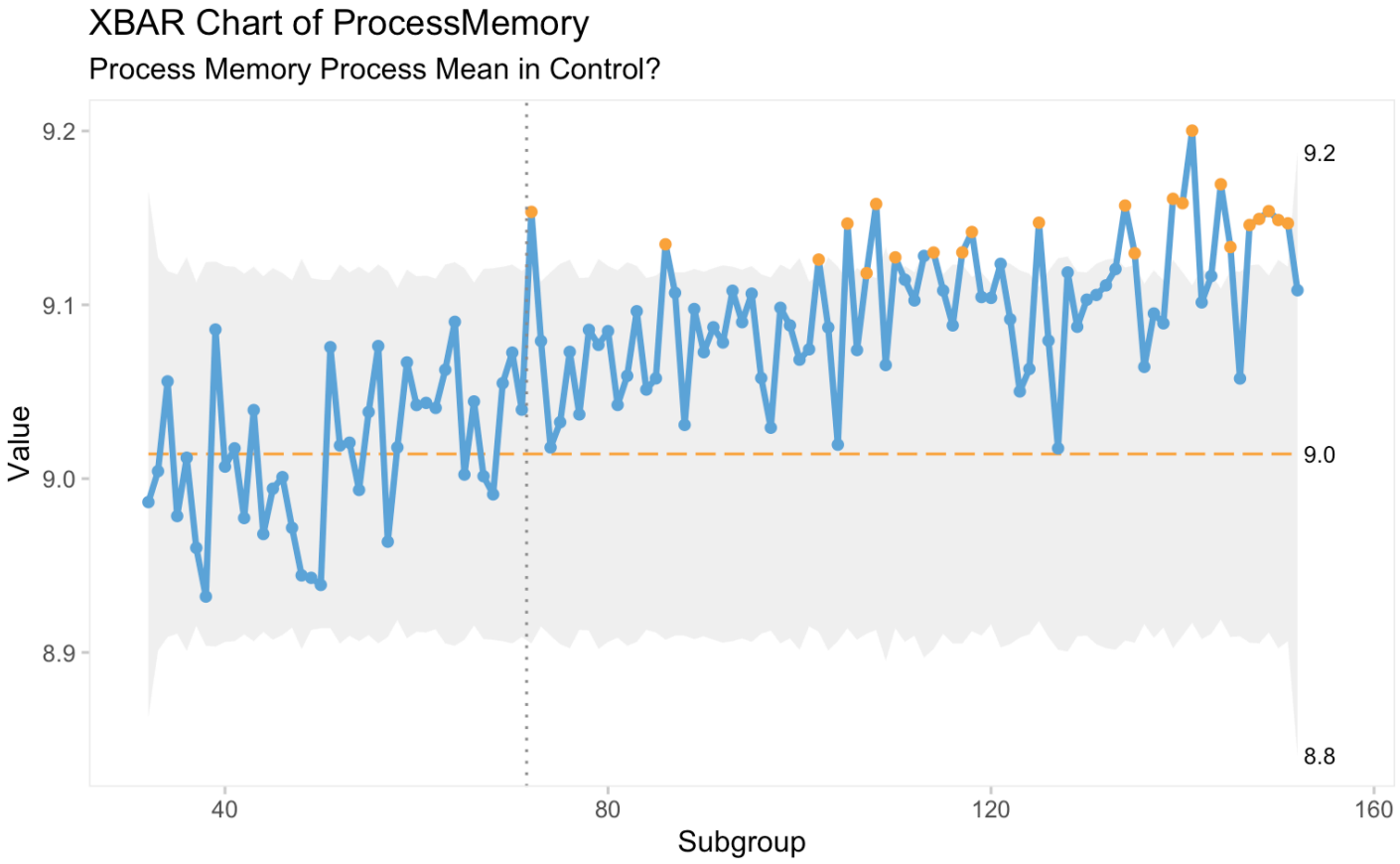
Creating control charts to answer questions.

a) Is the memory usage of the server in control?

First, we will look at the 'xbar' chart which can plot the means of the sets of consecutive points, and shows whether the process mean is in control.

Hide

```
#xbar for mean
mean_memory_p <- qicharts2::qic(x = day_of_year, y = ProcessMemory, data = monitor
_data, chart = "xbar", subtitle = "Process Memory Process Mean in Control?", freez
e = 40)
plot(mean_memory_p)
```



Hide

```
mean_memory_summary <- summary(mean_memory_p)
mean_memory_summary
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	84	10	17

1 row | 1-10 of 16 columns

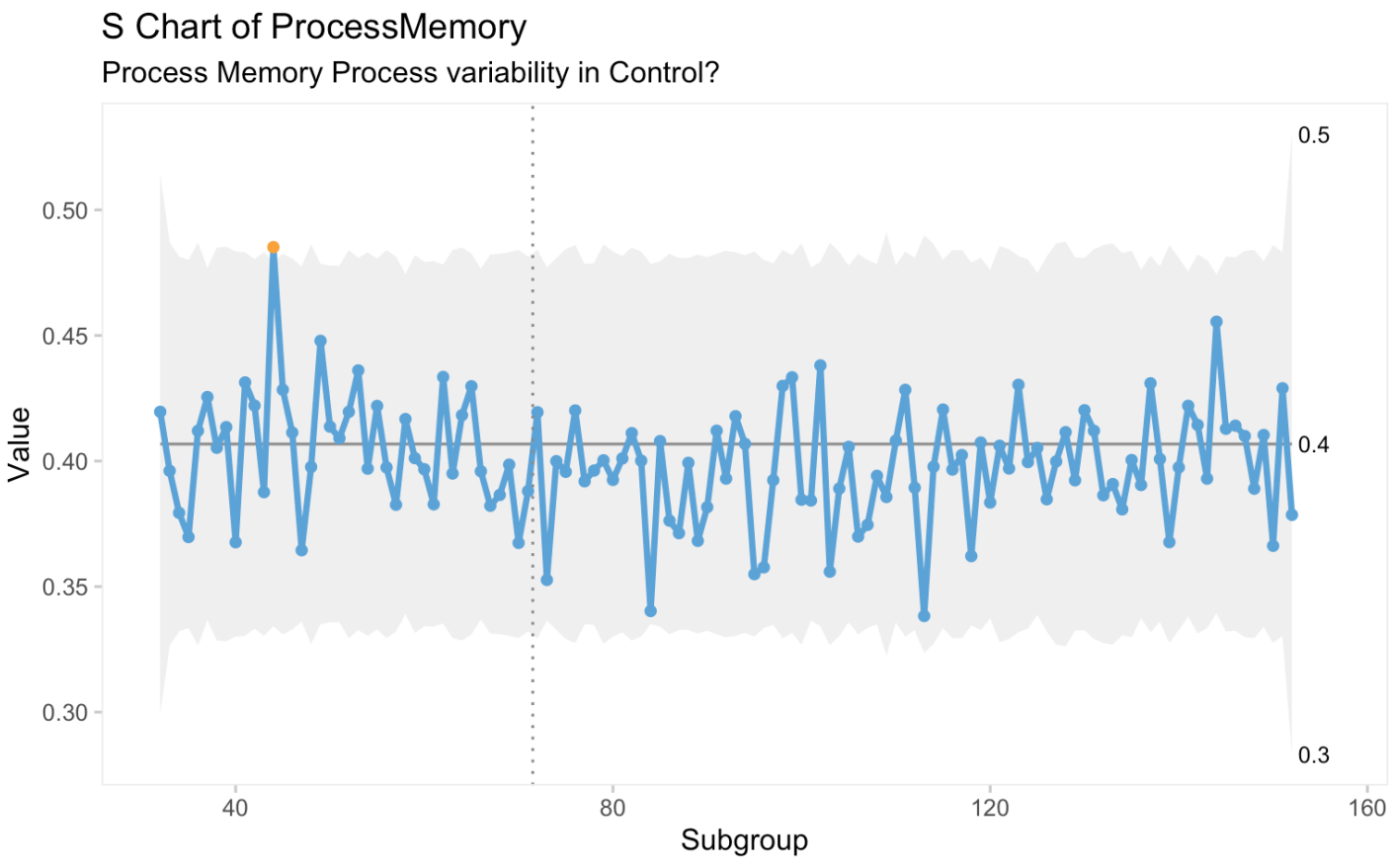
The sigma signal is 23 and the runs signal is 1. The sigma signal of 23 indicates that there were 23 occurrences of the line going beyond the grey area of the plot. The runs signal of 1 also indicates that points were above (or below, but not in this case) the central (expected) line for a sequence of points. Specifically, runs.signal indicates that longest.run was greater than longest.run.max in this case. Ultimately, this result tells us that the memory usage of the server is not in control.

Next, we will look at the standard deviation of consecutive points using the 's' chart. The 's' chart shows whether the process variablity is in control.

Plotting standard deviation of sets of consecutive points using S chart. An S chart shows whether the process variability is in control.

Hide

```
sd_memory_p <- qicharts2::qic(x = day_of_year, y = ProcessMemory, data = monitor_d
ata, chart = "s", subtitle = "Process Memory Process variability in Control?", fre
eze = 40)
plot(sd_memory_p)
```



Hide

```
sd_memory_summary <- summary(sd_memory_p)
sd_memory_summary
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	7	10	57

1 row | 1-10 of 16 columns

Hide

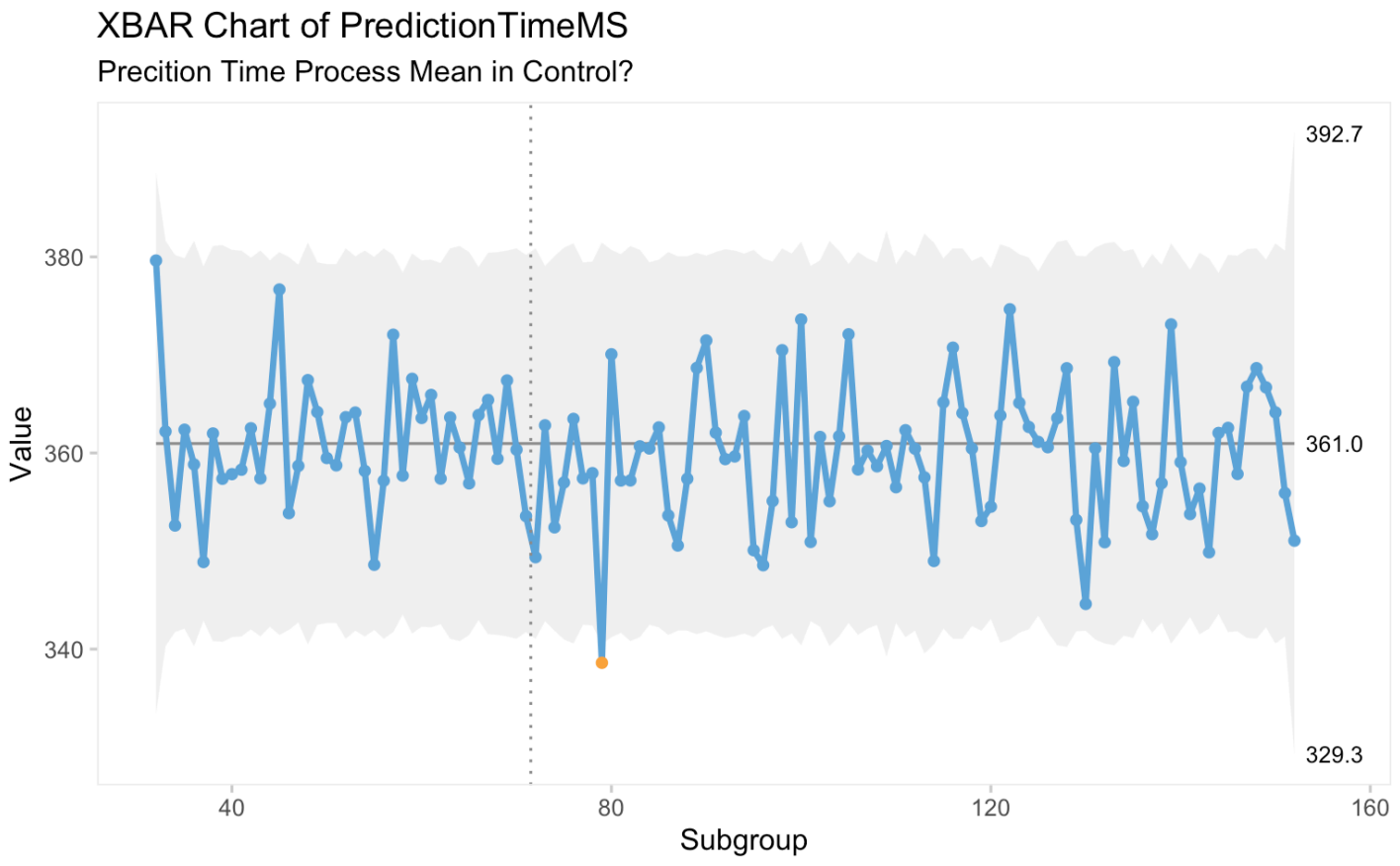
```
NA
```

The sigma signal is 1 and runs signal is 0. This suggests there was one time where the line went beyond the limit, also visualised in the graph. The longest.run was less than the longest.run.max (7< 10) in this instance, but n.crossings was greater than n.crossings.min (57 > 51). This tells us that the process memory variability is not in control.

b) Is the prediction time of the model in control?

Hide

```
mean_predictionT_p <- qicharts2::qic(x = day_of_year, y = PredictionTimeMS, data =
monitor_data, chart = "xbar", subtitle = "Precition Time Process Mean in Control?"
, freeze = 40)
plot(mean_predictionT_p)
```



Hide

```
mean_pred_time_sumamry <- summary(mean_predictionT_p)
mean_pred_time_sumamry
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	5	10	61

1 row | 1-10 of 16 columns

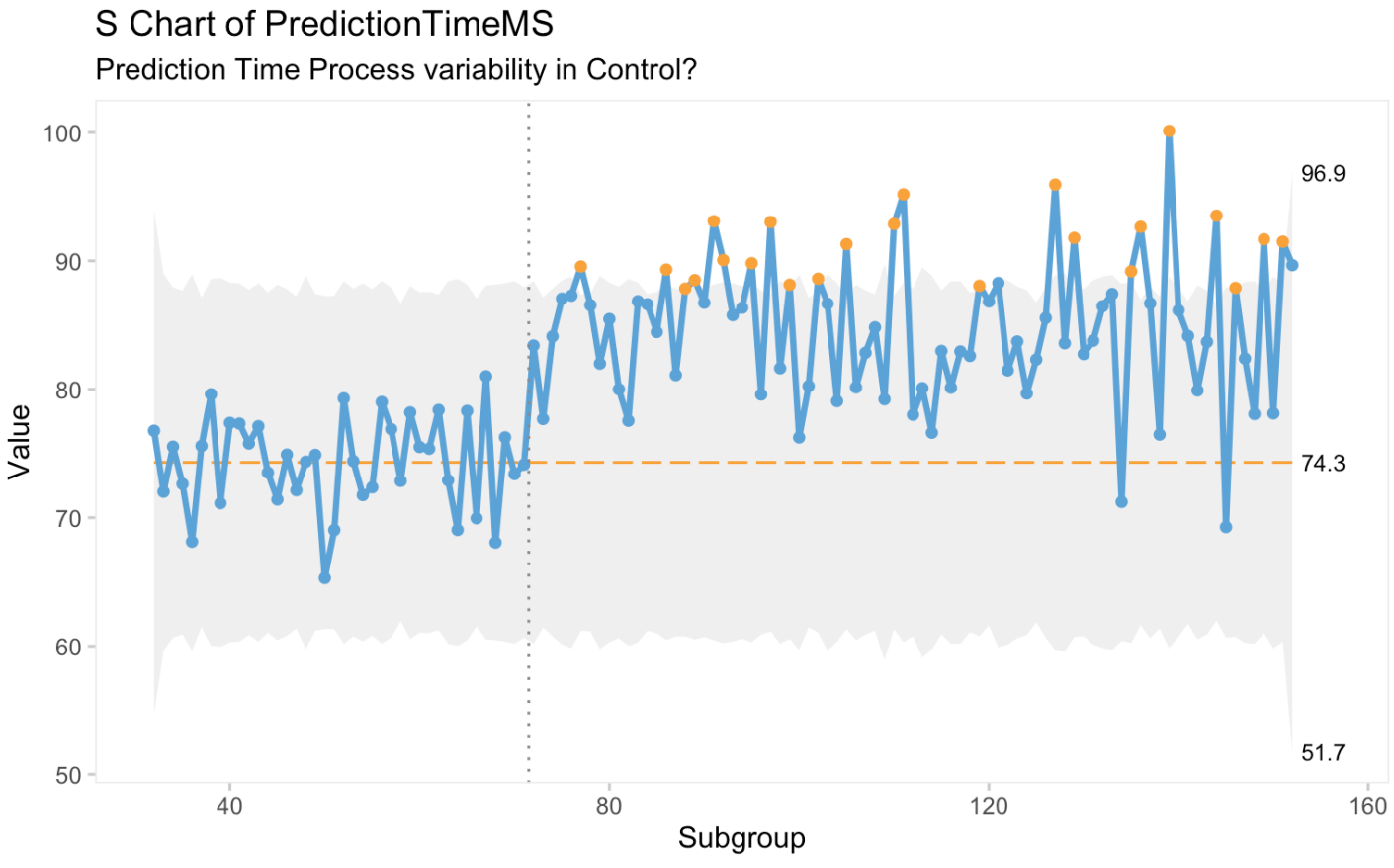
Hide

```
NA
```

The results above indicate that the prediction time process mean is not in control, as the sigma signal is a value of 1 (the line has gone beyond the limit once). Signal runs is 0, as Longest.run is less than longest.run.max (5 < 10), but n.crossings is greater than n.crossings.min (61 > 51).

Hide

```
sd_predictionT_p <- qicharts2::qic(x = day_of_year, y = PredictionTimeMS, data = m
onitor_data, chart = "s", subtitle = "Prediction Time Process variability in Contr
ol?", freeze = 40)
plot(sd_predictionT_p)
```



Hide

Hide

```
sd_pred_time_summary <- summary(sd_predictionT_p)
sd_pred_time_summary
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	62	10	28

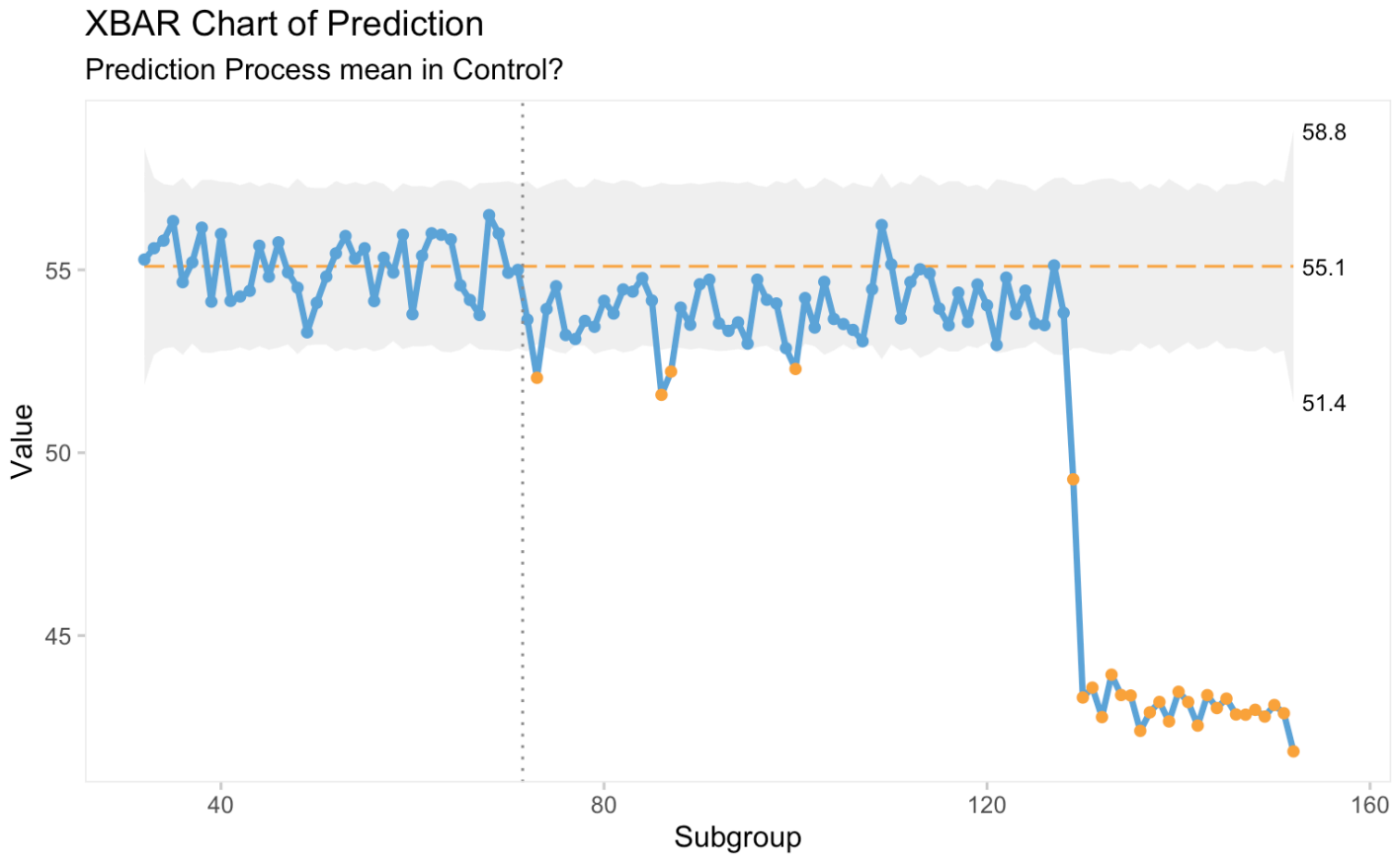
1 row | 1-10 of 16 columns

The results above show that the prediction time process variability is also not in control. The sigmal signal of 23 indicates there were 23 instances of the line crossing the grey area boundary. The runs signal of 1 also shows this. Longest.run is greater than longest.run.max, but n.crossings is less than n.crossings.min. Ultimately, the prediction time process is not in control.

c) Is the stream of predictions in control?

Hide

```
mean_prediction_p <- qicharts2::qic(x = day_of_year, y = Prediction, data = monito
r_data, chart = "xbar", subtitle = "Prediction Process mean in Control?", freeze =
40)
plot(mean_prediction_p)
```



Hide

```
mean_predicition_summary <- summary(mean_prediction_p)
mean_predicition_summary
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	39	10	23

1 row | 1-10 of 16 columns

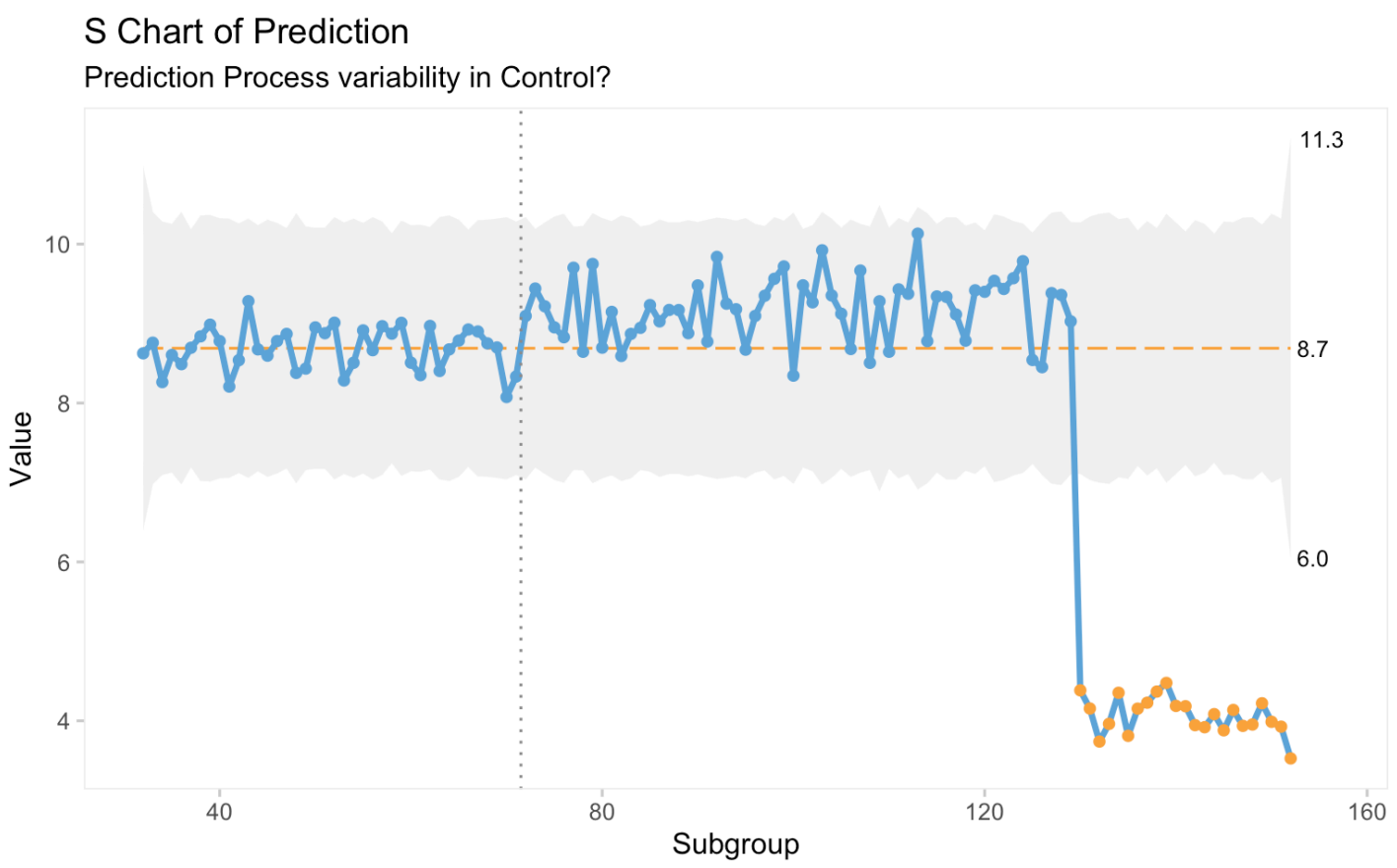
Hide

```
NA
```

The above results show that the stream of predictions process is not in control. runs.signal is 1, indicating there were instances where the model is not in control. Sigma.signal confirms there were 28 times the model crossed the threshold limit. Longest.run is greater than longest.run.max and n.crossings is less than n.crossings.min.

Hide

```
sd_prediction_p <- qicharts2::qic(x = day_of_year, y = Prediction, data = monitor_
data, chart = "s", subtitle = "Prediction Process variability in Control?", freeze
= 40)
plot(sd_prediction_p)
```



Hide

```
sd_prediction_summary <- summary(sd_prediction_p)
sd_prediction_summary
```

facet1 <dbl>	facet2 <dbl>	p... <int>	n.o... <int>	n.useful <int>	longest.run <int>	longest.run.max <dbl>	n.crossings <dbl>	n.cro
1	1	1	1	121	121	23	10	36

1 row | 1-10 of 16 columns

The S chart shows the variability of control for the prediction process. This is not in control, observed by runs.signal being 1, indicating the limit has been crossed. Sigma.signal confirms that there were 23 instances where the limit was breached. Longest.run is greater than longest.run.max and n.crossings is less than n.crossings.min.

Summarising results in a table:

Measurement <chr>	Xbar_Runs_Signal <int>	Xbar_Breaches <int>	S_Runs_Signal <int>	S_Breaches <int>	Overall <chr>
Memory	1	23	0	1	Out of c
Prediction Time	0	1	1	23	Out of c

Prediction	1	28	1	23	Out of c

3 rows | 1-6 of 7 columns

By summarising these results in a table, we can observe clearly that Memory, Prediction time and Prediction were not in control, both in terms of process mean and variability.