Understanding Data over time
Run Charts & Control Charts

Lothian Analytical Services

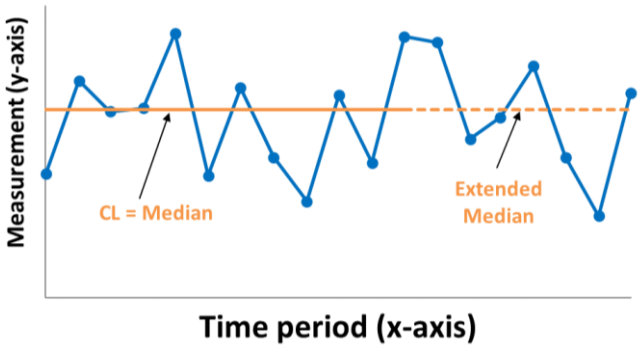
LothianQuality
BETTER HEALTH, BETTER CARE, BETTER VALUE

Run and Control charts are some of the tools that we can use to understand the variation within a data set. Run charts are a simple tool that is used frequently in quality improvement projects they are helpful when creating charts by hand with small amounts of data often collected manually. Control Charts require more data than a run chart to create and are more likely to be used over a longer time frame and can also be used to monitor processes and outcomes.
When constructing Dashboards in Tableau the default will be to use the appropriate Control Chart, though run charts may be used when there is less than 16 data points available or as part of early development.

Run Charts

- Run Charts are used to assess the type of variation that exists in a system of interest and whether it is random or non-random.

- A measure is assessed in order, most often as time series on the x-axis, a median is used as a centreline and 4 rules are used to test the data. These are:



1. A **shift**: six or more consecutive data points either all above or all below the median. Points on the median do not count towards or break a shift.
2. A **trend**: five or more consecutive data points that are either all increasing or decreasing in value. If two points are the same value ignore one when counting.
3. **Too many or too few runs**: If there are too many or too few runs (i.e. the median is crossed too many or too few times) that's a sign of non-random variation. You need to look up a statistical table to see what an appropriate number of runs to expect would be.
4. An **astronomical data point**: a data point that is clearly different from all others. This relies on judgement. Every data set has a highest and lowest point, but these won't necessarily be astronomical data points. Different people looking at the same graph would be expected to recognise the same data point as astronomical (or not).

Provost & Murray, Healthcare Data Guide 2011

Fixing and recalculating the median

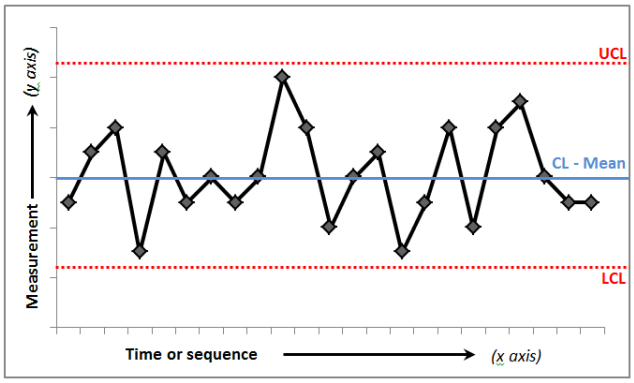
- The period over which the median is calculated should be fixed and extended to better detect presence of non-random variation as new data is added to the existing data set.
- It is recognised that to do this judgement is normally required and expert QI measurement advice should be sought if needed.
- Current defaults for fixing/extending a Run Chart's median are outlined in the following table:

1. **Points for a baseline median**: for data series with less than 12 points, a median can be applied across the entire duration and updated as each new point becomes available. Once at 12 points, the baseline median should be fixed and "extended" forward.
2. **Recalculating the median**: When a shift is detected in the extended zone (6 points) and sustained (a further 3 points) a new median can be calculated. The recalculated median will initially be based on nine points and considered temporary – recalculate to include the 12th point when available and then the median can be fixed again Note that by this point there should be at least 21 data points and enough to consider producing an SPC chart with temporary limits.

https://www.healthcareimprovementscotland.org/previous_resources/misc/guide_to_using_run_charts.aspx

SPC Charts

- Statistical Process Control Charts or SPC Charts (also called Shewhart Control Charts) are also used to assess the type of variation that exists in a system of interest.
- SPC Charts use control limits in addition to a centreline to allow more tests of variation as well as helping to determine the capability of the system.



- In SPC Charts we test for the presence of Special Cause variation using a variety of tests. The following table provides a summary of these rules:

1. **Astronomical point - 1 point outside the +/- 3 sigma limits** - A point exactly on a control limit is not considered outside the limit.
2. **Shift - 8 successive or more consecutive points above (or below) the centreline** - A point exactly on the centreline does not cancel or count towards a shift.
3. **Trend - 6 or more consecutive points steadily increasing or decreasing** - Two consecutive points of the same value do not cancel or add to a trend.
4. **2/3 rule - 2 out of 3 successive points near a control limit (outer one third)** - Two out of three consecutive data points located close to one of the control limits (within 2 and 3 sigma).
5. **15 points rule - 15 consecutive points within 1 sigma on either side of the centreline** - This is known as “hugging the centreline”.

Provost & Murray, Healthcare Data Guide 2011

Centre lines/control limits:

- The period over which the centreline and control limits are calculated, should be fixed and extended to better detect presence of special cause variation as new data is added to the data set.
- It is recognised that to do this judgement is normally required and expert QI measurement advice should be sought if needed.
- Current standards for fixing/extending SPC Charts’ mean and control limits (suitable for Tableau) are outlined in the following table:

1. **Points for setting mean and control limits**: 24 data points should be used to calculate centreline and control limits. Trial limits can be used from 16 points (use a run chart for less) but recalculated with each new point until 24 are available.
2. **Recalculating mean and control limits**: Control limits are normally recalculated when evidence of change is clear, and the change is ongoing, understood and has occurred for long enough for representative control limits to be calculated. This normally requires some level of judgement but for the purposes of a dashboard environment we will recalculate limits when there has been a shift of 8 points which has been sustained for a further 8 points, at this stage (16 points) new limits will be calculated and considered temporary until 24 data points are available.

Selecting SPC Charts

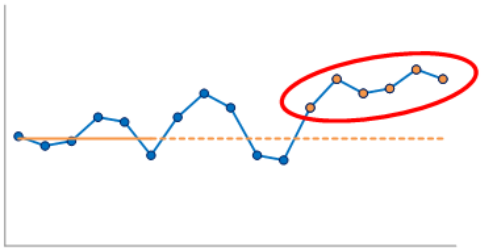
- A number of alternative types of SPC charts are used depending on the data being visualised (P, XMR, U, XbarS, C). The most appropriate chart for the data should be used.
- If there is uncertainty about chart choice, XmR charts could be used until use of the most appropriate chart type is feasible.

Appendices

Run chart Rules

Rule 1: a shift

6 or more consecutive data points above or below the median



Rule 2: a trend

5 or more consecutive data points all going in same direction

Rule 3: too many or too few runs*

This chart has 24 useful data points and it has 7 runs, which is less than the minimum in the table

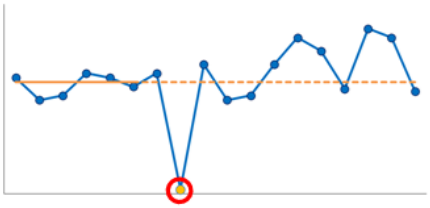
This chart has 24 useful data points and it has 19 runs, which is more than the maximum in the table

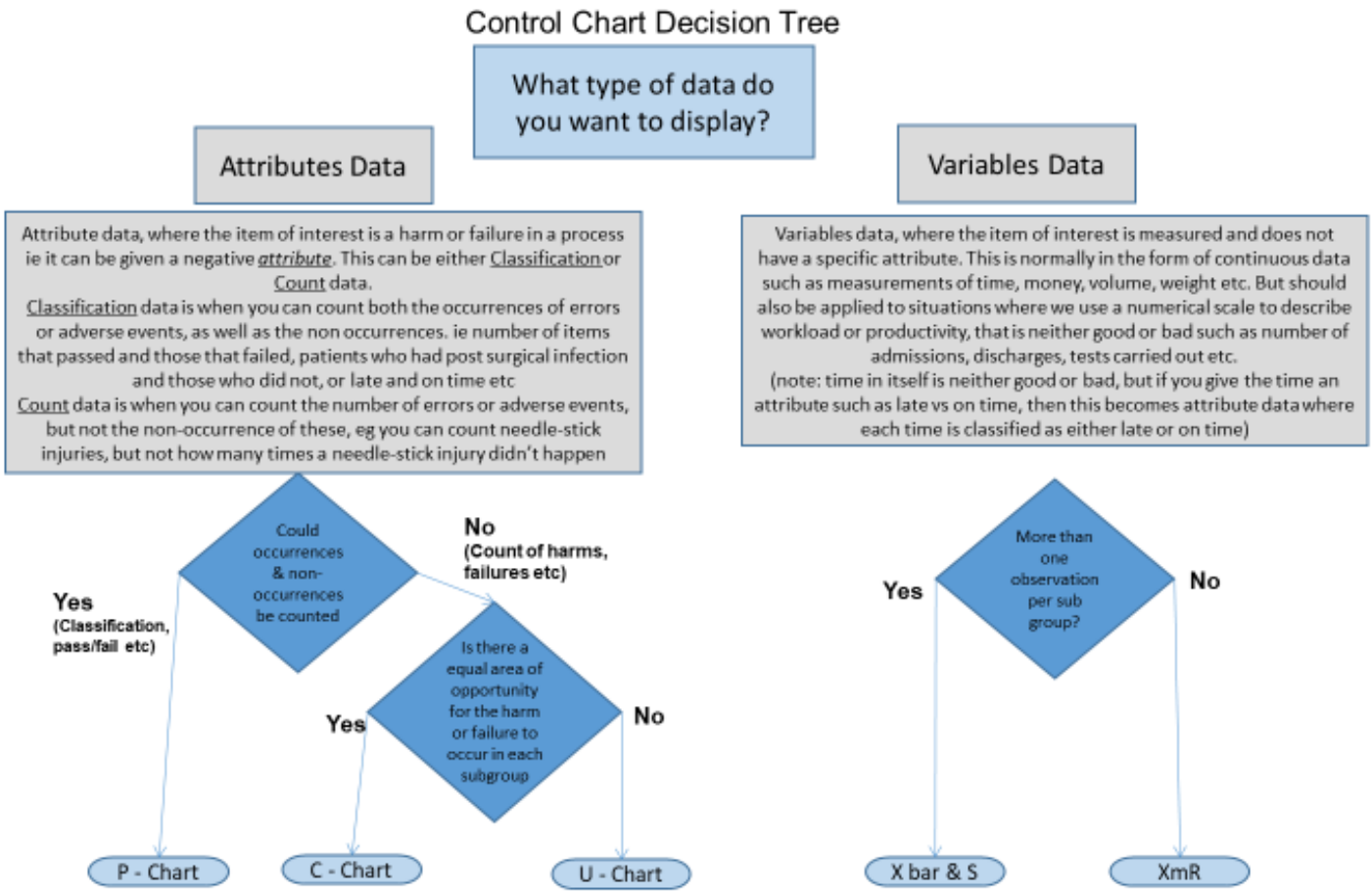
	Useful data points	Minimum number of runs	Maximum number of runs
Number of useful data points is the number of points that are not on the median. The corresponding numbers in the table are the minimum and maximum number of runs that if exceeded are a signal of non random variation	15	5	12
	16	5	13
	17	5	13
	18	6	14
	19	6	15
	20	6	16
	21	7	16
	22	7	17
	23	7	17
	24	8	18
	25	8	18
	26	9	19
	27	10	19
	28	10	20
	29	10	20
	30	11	21

*A run is a group of consecutive points on the same side of the median

Rule 4: astronomical point

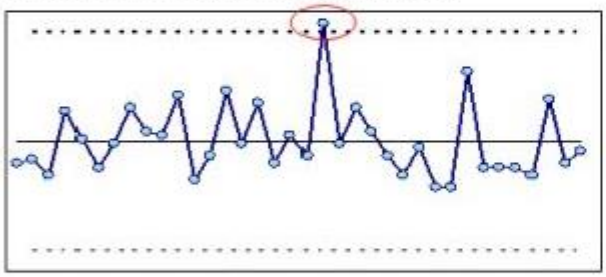
A point which is obviously very different from all other points



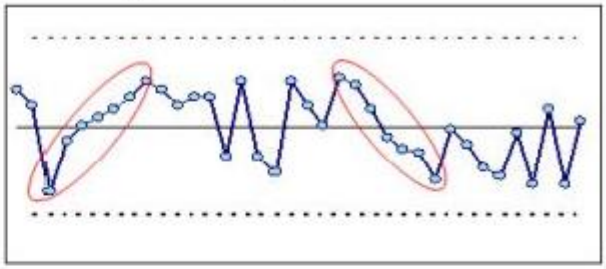

Control Chart Decision Tree
What type of data do you want to display?
Attributes Data
Variables Data
Attribute data, where the item of interest is a harm or failure in a process ie it can be given a negative attribute. This can be either Classification or Count data.
Classification data is when you can count both the occurrences of errors or adverse events, as well as the non occurrences. ie number of items that passed and those that failed, patients who had post surgical infection and those who did not, or late and on time etc
Count data is when you can count the number of errors or adverse events, but not the non-occurrence of these, eg you can count needle-stick injuries, but not how many times a needle-stick injury didn't happen
Variables data, where the item of interest is measured and does not have a specific attribute. This is normally in the form of continuous data such as measurements of time, money, volume, weight etc. But should also be applied to situations where we use a numerical scale to describe workload or productivity, that is neither good or bad such as number of admissions, discharges, tests carried out etc.
(note: time in itself is neither good or bad, but if you give the time an attribute such as late vs on time, then this becomes attribute data where each time is classified as either late or on time)
Could occurrences & non-occurrences be counted
Yes
(Classification, pass/fail etc)
No
(Count of harms, failures etc)
Is there a equal area of opportunity for the harm or failure to occur in each subgroup
Yes
No
P - Chart
C - Chart
U - Chart
More than one observation per sub group?
Yes
No
X bar & S
XmR

API Rules for Detecting Special Cause

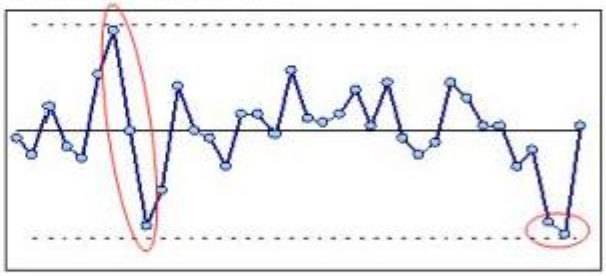
A single point outside the control limits



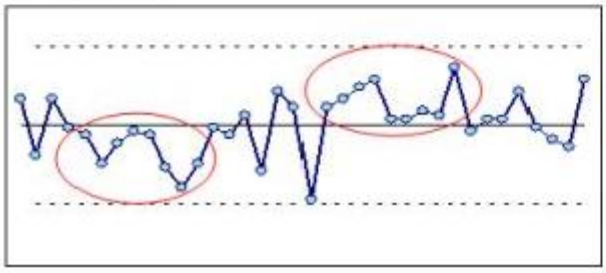
Six consecutive points increasing (trend up) or decreasing (trend down)



Two out of three consecutive points near a control limit (outer one-third)



Eight or more consecutive points above or below the centerline



Fifteen consecutive points close to the centerline (inner one-third)