Exploration Data Analysis Report

Team Members

- Isaac Adoboe - 202384695
- Prince Afari - 202485668
- Mary Ntim - 202398287
- Rahmat Efua Etuaful -

```
## Package 'qcc' version 2.7

## Type 'citation("qcc")' for citing this R package in publications.

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Statistical Process Control Analysis of Brewery Operations

This report presents a Statistical Process Control (SPC) analysis of brewery operations based on data from 100 brewing batches across 8 beer styles and 10 production locations. The analysis focuses on identifying process variations, control patterns, quality correlations, and defect rates to optimize brewing operations.

Introduction

Statistical Process Control (SPC) is a method for monitoring, controlling, and improving a process through statistical analysis. In brewing operations, SPC helps maintain product consistency, identify sources of variation, and prevent quality issues before they reach customers.

This analysis examines brewery data using SPC principles to understand:

1. Process stability and capability
2. Factors affecting quality and defect rates
3. Variations across beer styles and production locations
4. Process capability against specification limits
5. Opportunities for process improvement

Data Overview

The dataset contains 100 brewing batches distributed across 8 beer styles and 10 production locations. For SPC analysis, we've applied several control chart methods:

1. **Variable Data Analysis**: X-bar and R charts with subgroups of size 5
2. **Attribute Data Analysis**:
    - p charts (proportion of defective units)
    - np charts (number of defective units)
    - u charts (defects per unit)
3. **Process Capability Analysis**: Evaluating Quality Score against specification limits (7-10)

Process Stability Analysis

Variable Control Charts (X-bar and R Charts)

Quality Score and other key process variables were analyzed using X-bar and R charts with subgroups of size 5. This approach:

- Groups data points to better detect shifts in process mean and variability
- Provides insight into both between-group variation (X-bar chart) and within-group variation (R chart)
- Enhances ability to detect smaller shifts compared to individual values charts

The X-bar and R charts for Quality Score revealed the following:

- The process mean is stable with no points beyond control limits on the X-bar chart
- The R chart shows consistent within-group variability for most subgroups
- The control limits establish an expected range for brewing quality
- Sorting by Beer Style before subgrouping ensures meaningful comparison within styles

X-bar Chart Control Limits

- **Center Line (CL)**:

$$\bar{\bar{X}} \quad \text{(Grand average)}$$

- **Upper Control Limit (UCL)**:

$$\bar{\bar{X}} + A_2\bar{R}$$

- **Lower Control Limit (LCL)**:

$$\bar{\bar{X}} - A_2\bar{R}$$

Where A_2 is a constant that depends on subgroup size.

R Chart Control Limits

- **Center Line (CL)**:

$$\bar{R} \quad \text{(Average range)}$$

- **Upper Control Limit (UCL)**:

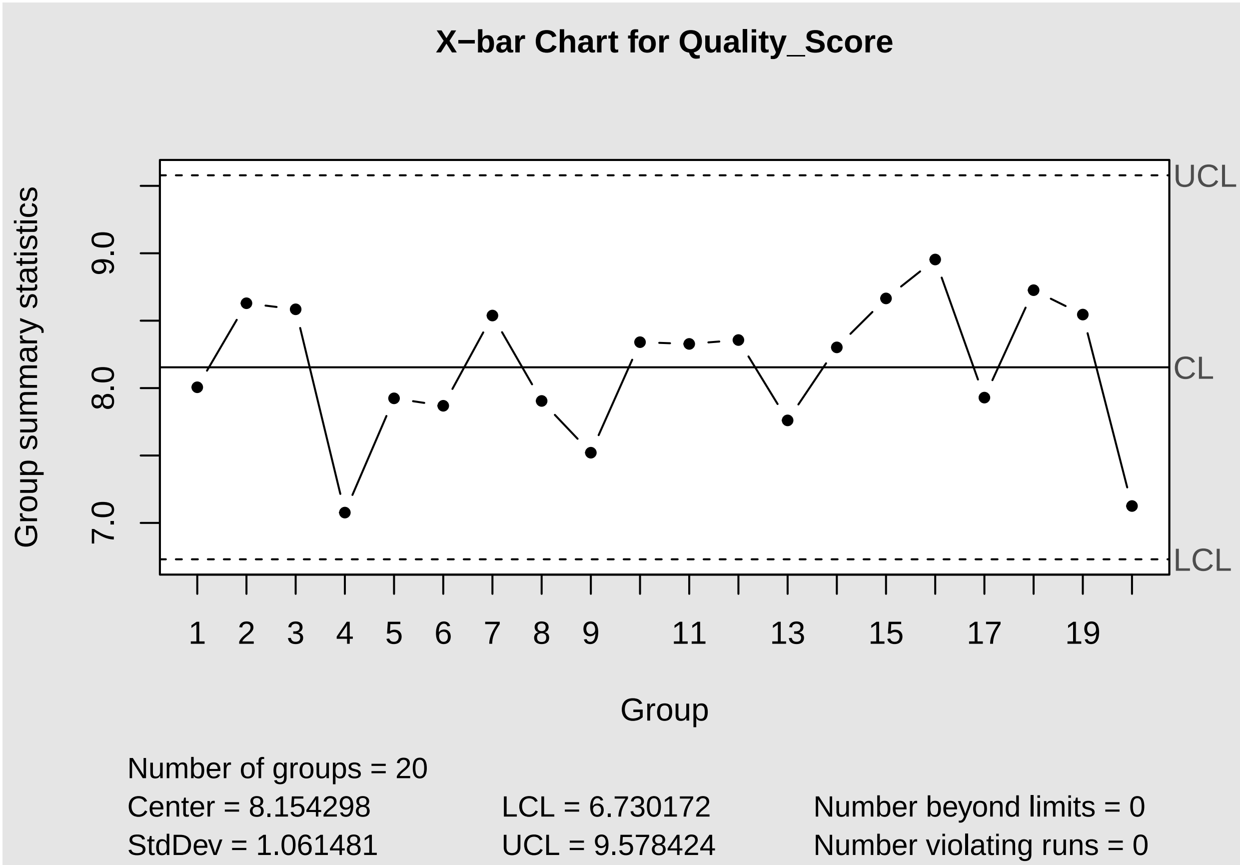
$$D_4\bar{R}$$

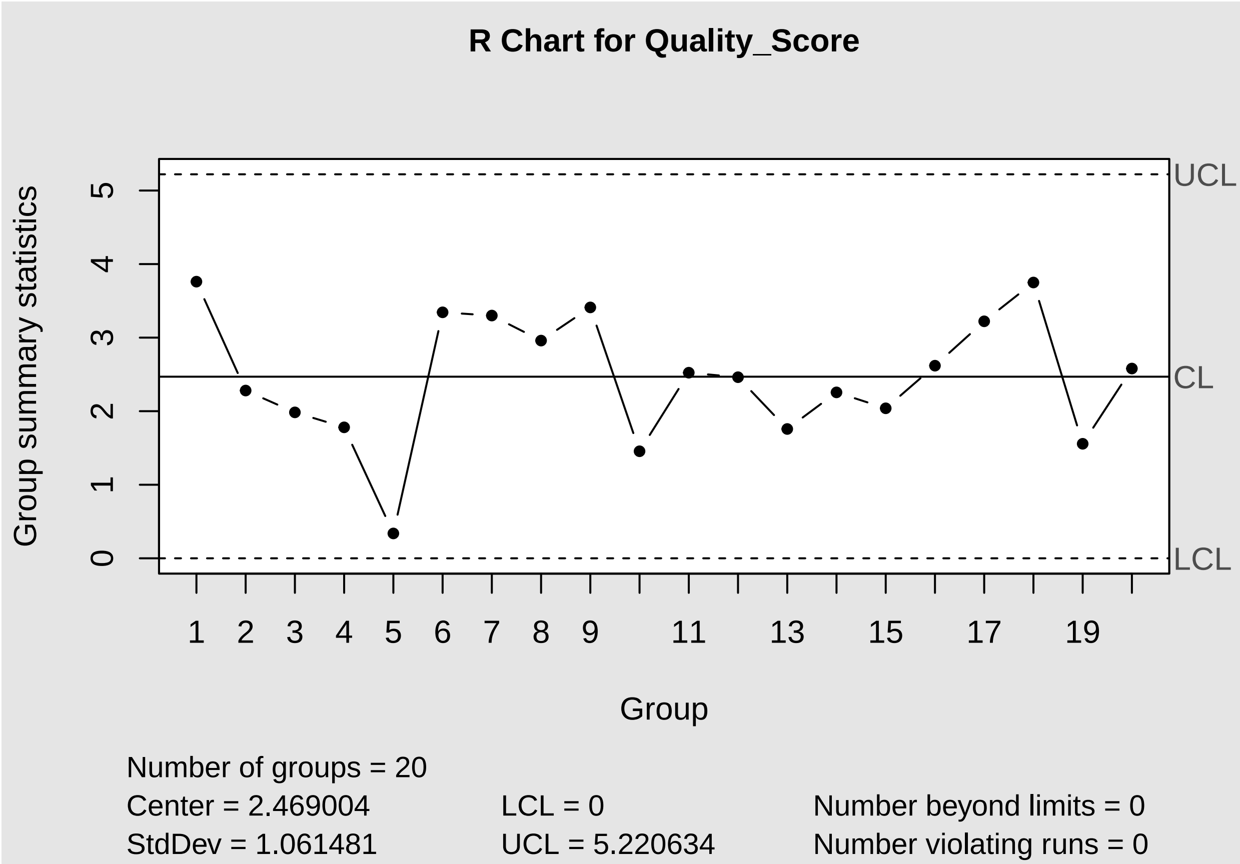
- **Lower Control Limit (LCL)**:

$$D_3\bar{R}$$

Where D_3 and D_4 are constants that depend on subgroup size.

```
## New names:
## Rows: 100 Columns: 16
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (2): Beer_Style, Location dbl (14): ...1, Fermentation_Time, Temperature,
## pH_Level, Gravity, Alcohol_C...
## i Use ‘spec()‘ to retrieve the full column specification for this data. i
## Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## * ‘‘ -> ‘...1‘
```


X−bar Chart for Quality_Score
Group summary statistics
7.0
8.0
9.0
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
11
13
15
17
19
Group
Number of groups = 20
Center = 8.154298
StdDev = 1.061481
LCL = 6.730172
UCL = 9.578424
Number beyond limits = 0
Number violating runs = 0



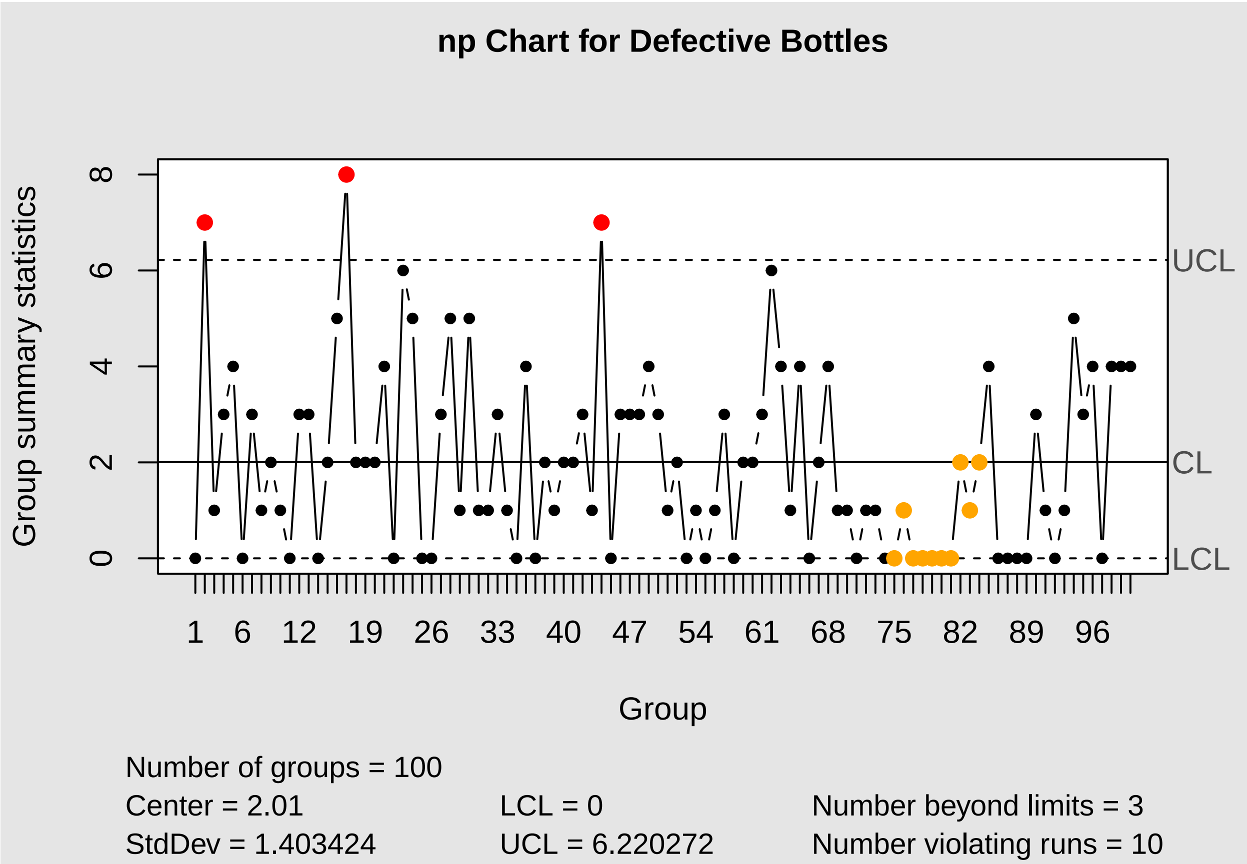

```
## List of 11
##  $ call      : language qcc(data = var_subgroups, type = "xbar", title = paste("X-bar Chart for",
##  $ type      : chr "xbar"
##  $ data.name : chr "var_subgroups"
##  $ data      : num [1:20, 1:5] 9.96 9.89 9.71 7.75 7.84 ...
##   ..- attr(*, "dimnames")=List of 2
##  $ statistics: Named num [1:20] 8.01 8.63 8.58 7.08 7.92 ...
##   ..- attr(*, "names")= chr [1:20] "1" "2" "3" "4" ...
##  $ sizes     : int [1:20] 5 5 5 5 5 5 5 5 5 5 ...
##  $ center    : num 8.15
##  $ std.dev   : num 1.06
##  $ nsigmas   : num 3
##  $ limits    : num [1, 1:2] 6.73 9.58
##   ..- attr(*, "dimnames")=List of 2
##  $ violations:List of 2
##  - attr(*, "class")= chr "qcc"
## List of 11
##  $ call      : language qcc(data = var_subgroups, type = "R", title = paste("R Chart for", var))
##  $ type      : chr "R"
##  $ data.name : chr "var_subgroups"
##  $ data      : num [1:20, 1:5] 9.96 9.89 9.71 7.75 7.84 ...
##   ..- attr(*, "dimnames")=List of 2
##  $ statistics: Named num [1:20] 3.761 2.281 1.984 1.781 0.338 ...
##   ..- attr(*, "names")= chr [1:20] "1" "2" "3" "4" ...
##  $ sizes     : int [1:20] 5 5 5 5 5 5 5 5 5 5 ...
##  $ center    : num 2.47
```

```
##   $ std.dev   : num 1.06
##   $ nsigmas   : num 3
##   $ limits    : num [1, 1:2] 0 5.22
##    ..- attr(*, "dimnames")=List of 2
##   $ violations:List of 2
##   - attr(*, "class")= chr "qcc"
```

Attribute Control Charts

Three types of attribute charts were used to analyze defect data:

1. **np Chart (Number of Defective Units)**:
   - Monitors the actual count of defective bottles per batch
   - Sensitive to changes in process quality affecting overall defect count



2. **p Chart (Proportion of Defective Units)**:
   - Tracks the percentage of defective bottles
   - Facilitates comparison across batches and beer styles
   - Used for style-specific analysis to identify quality differences

p Chart Control Limits

- **Center Line (CL)**:

$$\bar{p} \quad \text{(Average proportion defective)}$$

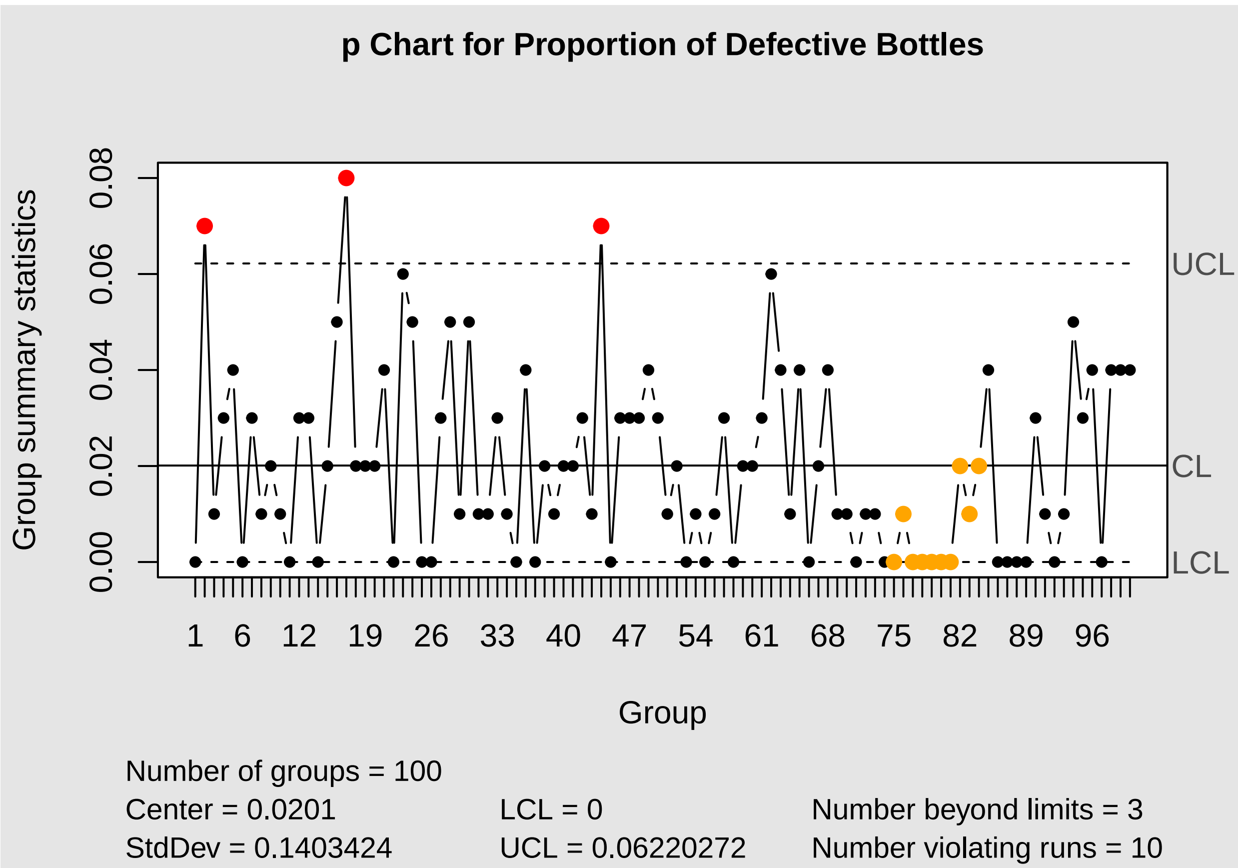
- **Upper Control Limit (UCL)**:

$$\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

- **Lower Control Limit (LCL)**:

$$\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Where n is the sample size.



3. **u Chart (Defects per Unit)**:
   - Counts multiple possible defects per bottle
   - More sensitive to subtle quality issues than p/np charts
   - Used for location-specific analysis to identify facility-related issues

u Chart Control Limits

- **Center Line (CL)**:

$$\bar{u} \quad \text{(Average defects per unit)}$$

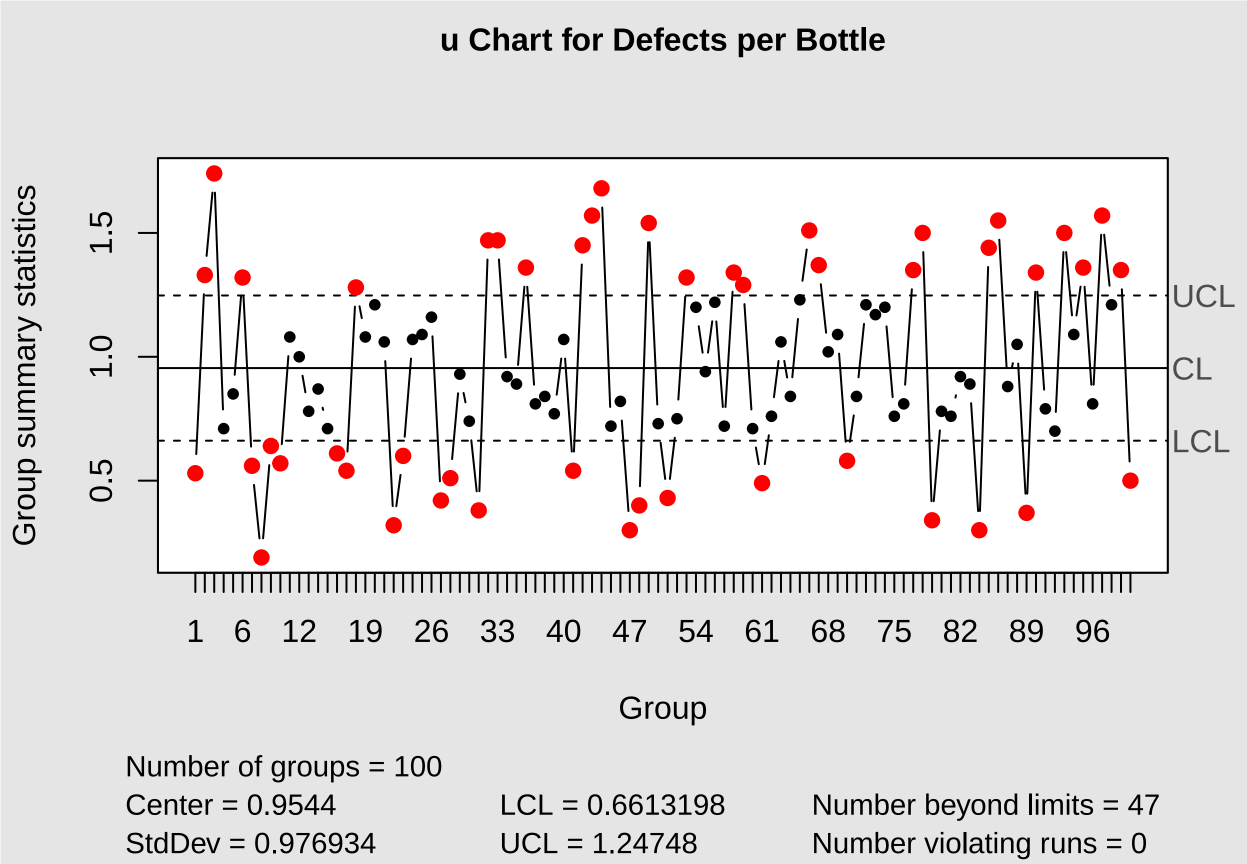
- **Upper Control Limit (UCL)**:

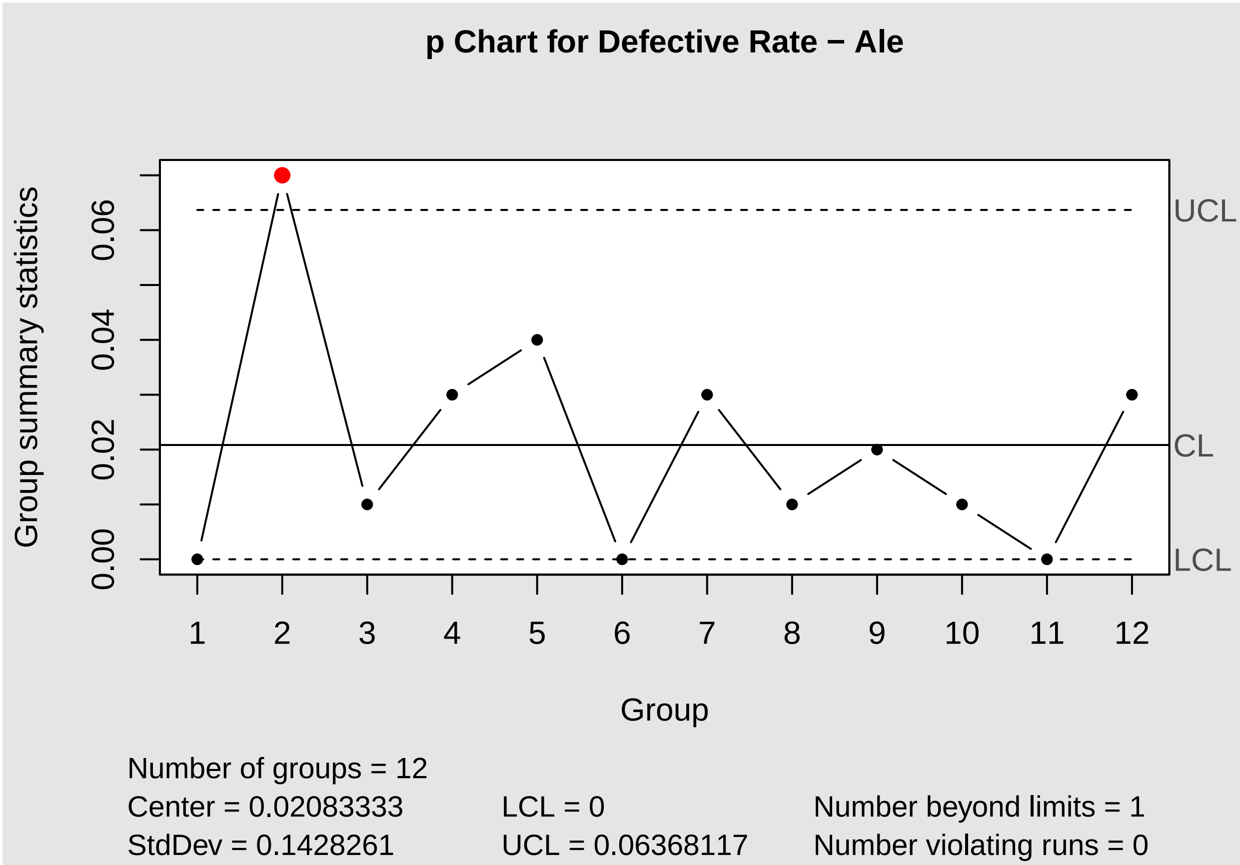
$$\bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$$

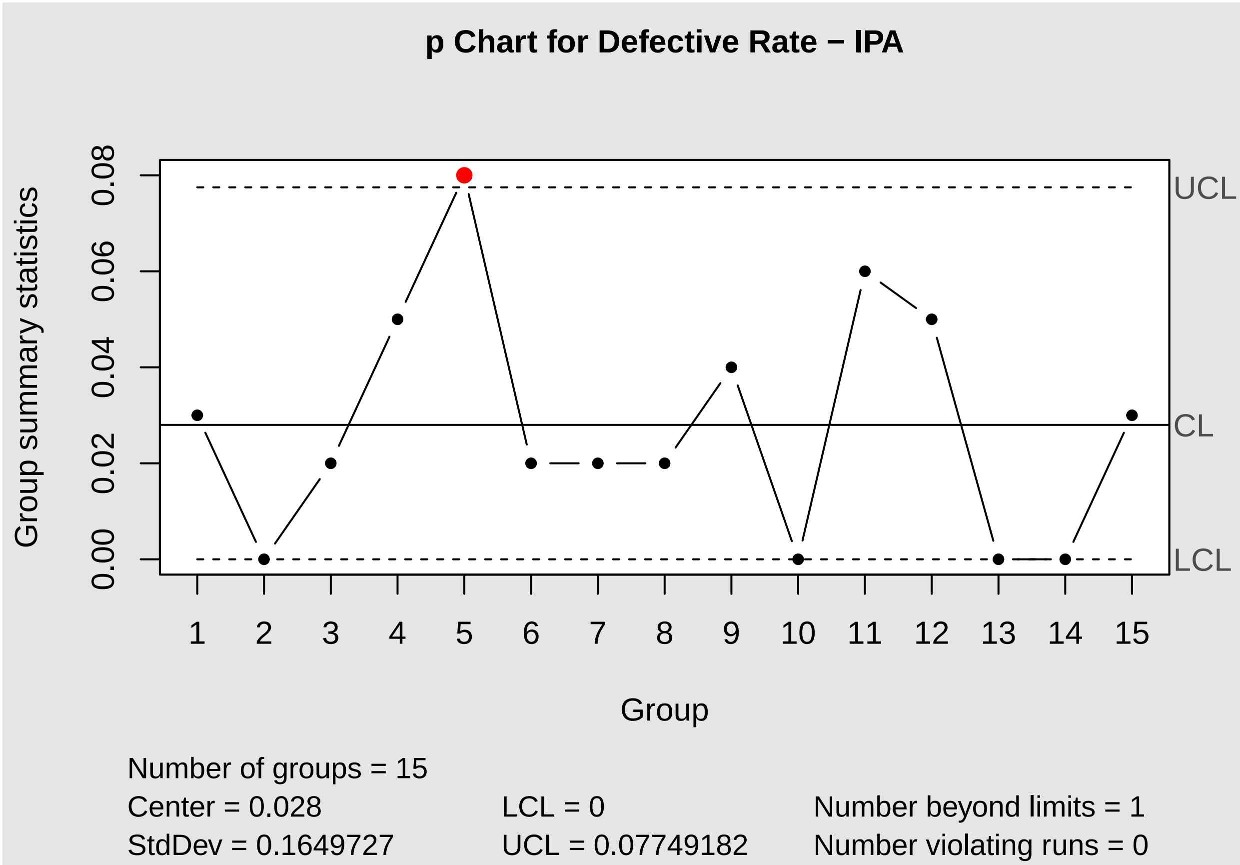
- **Lower Control Limit (LCL)**:

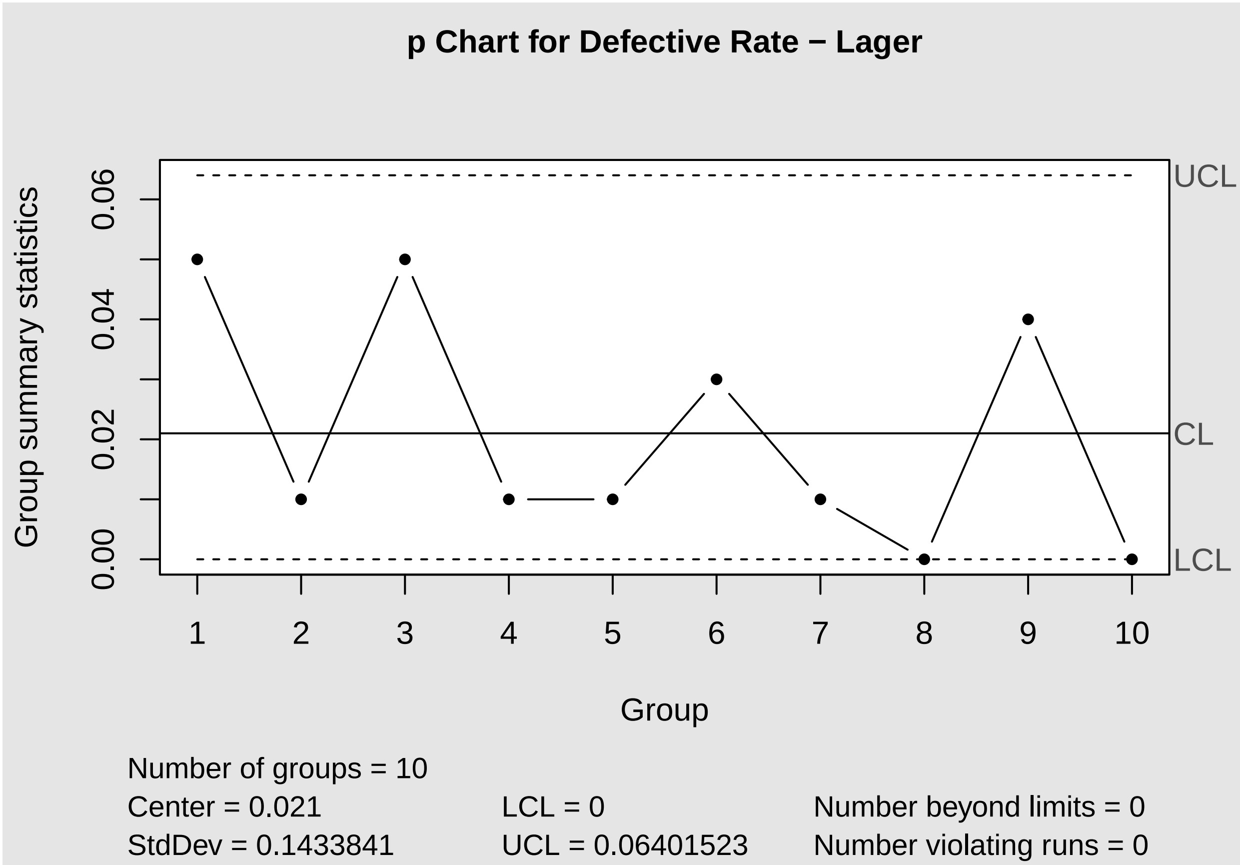
$$\bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$$

Where n is the sample size.

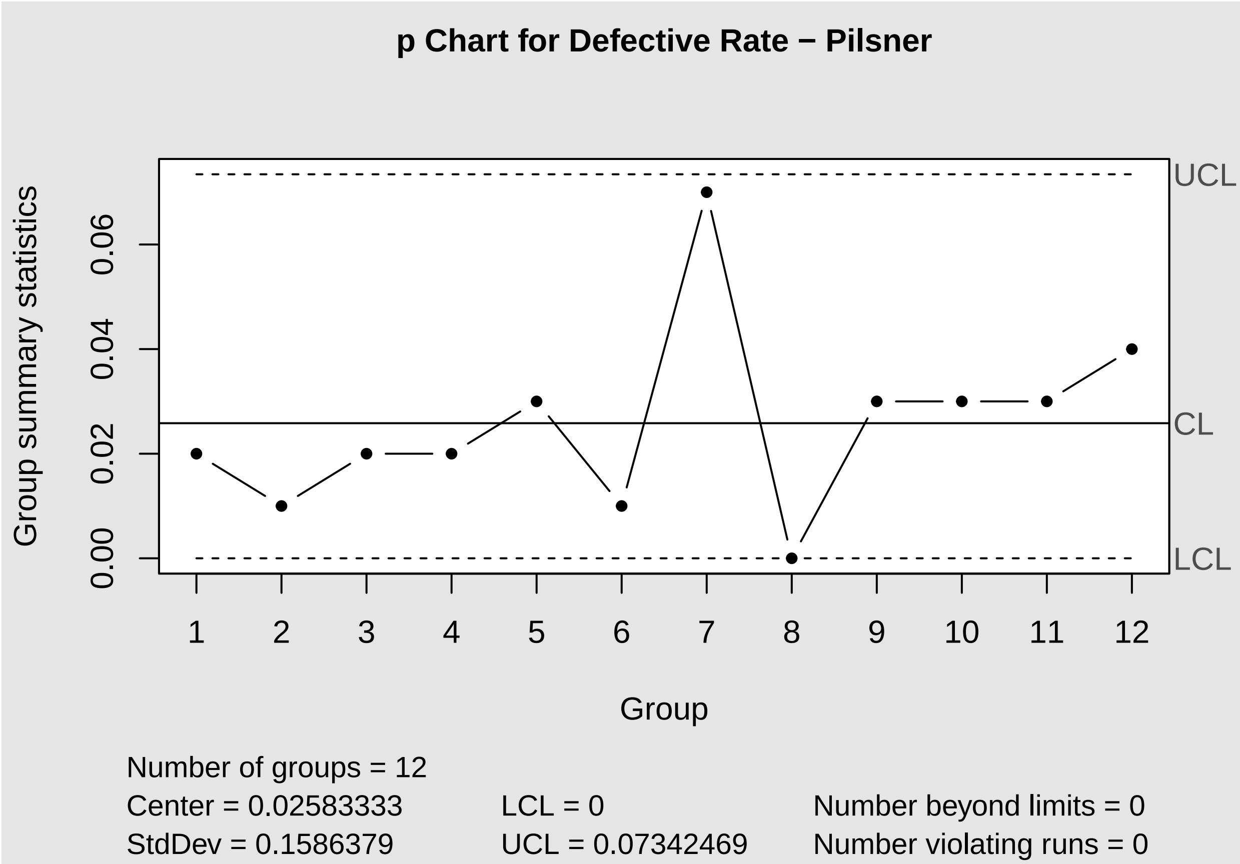


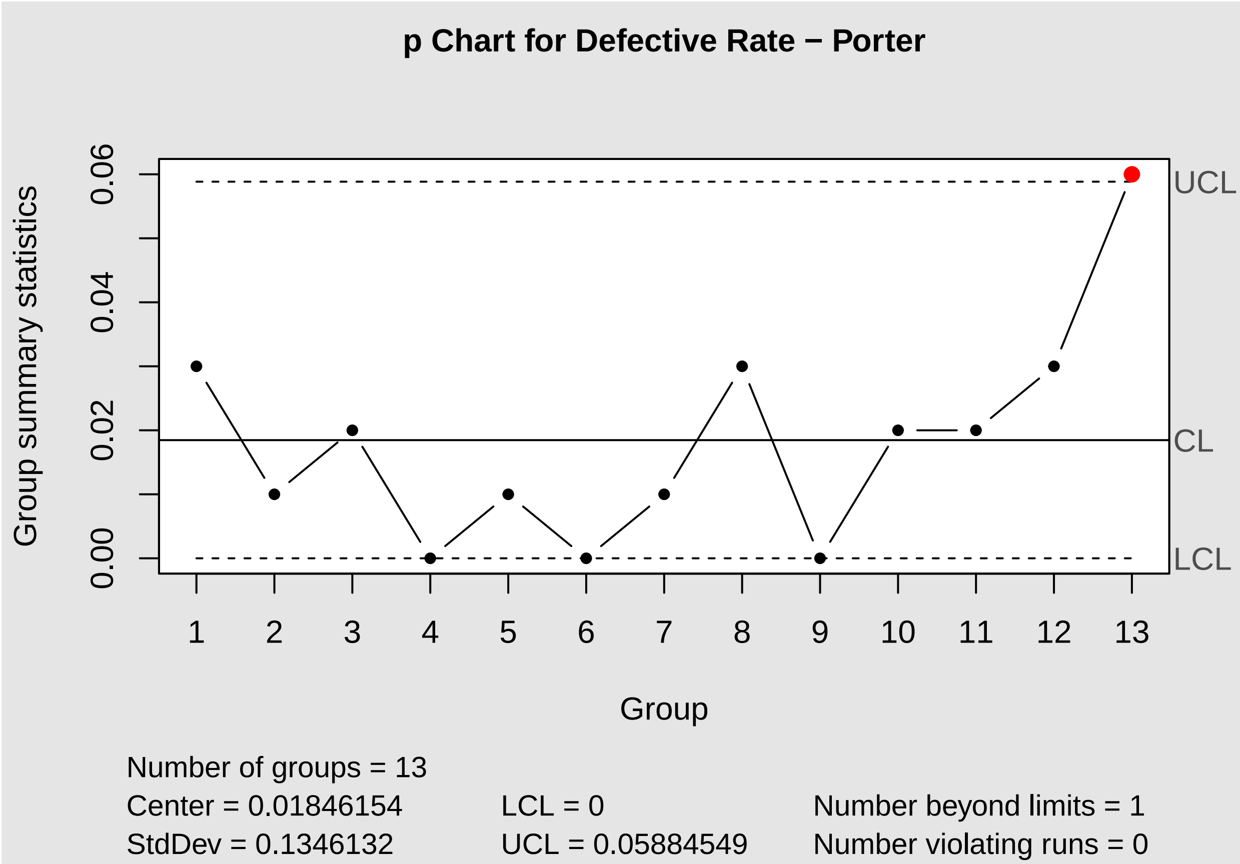

p Chart for Defective Rate − Ale
Group summary statistics
0.00
0.02
0.04
0.06
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
Group
Number of groups = 12
Center = 0.02083333
StdDev = 0.1428261
LCL = 0
UCL = 0.06368117
Number beyond limits = 1
Number violating runs = 0

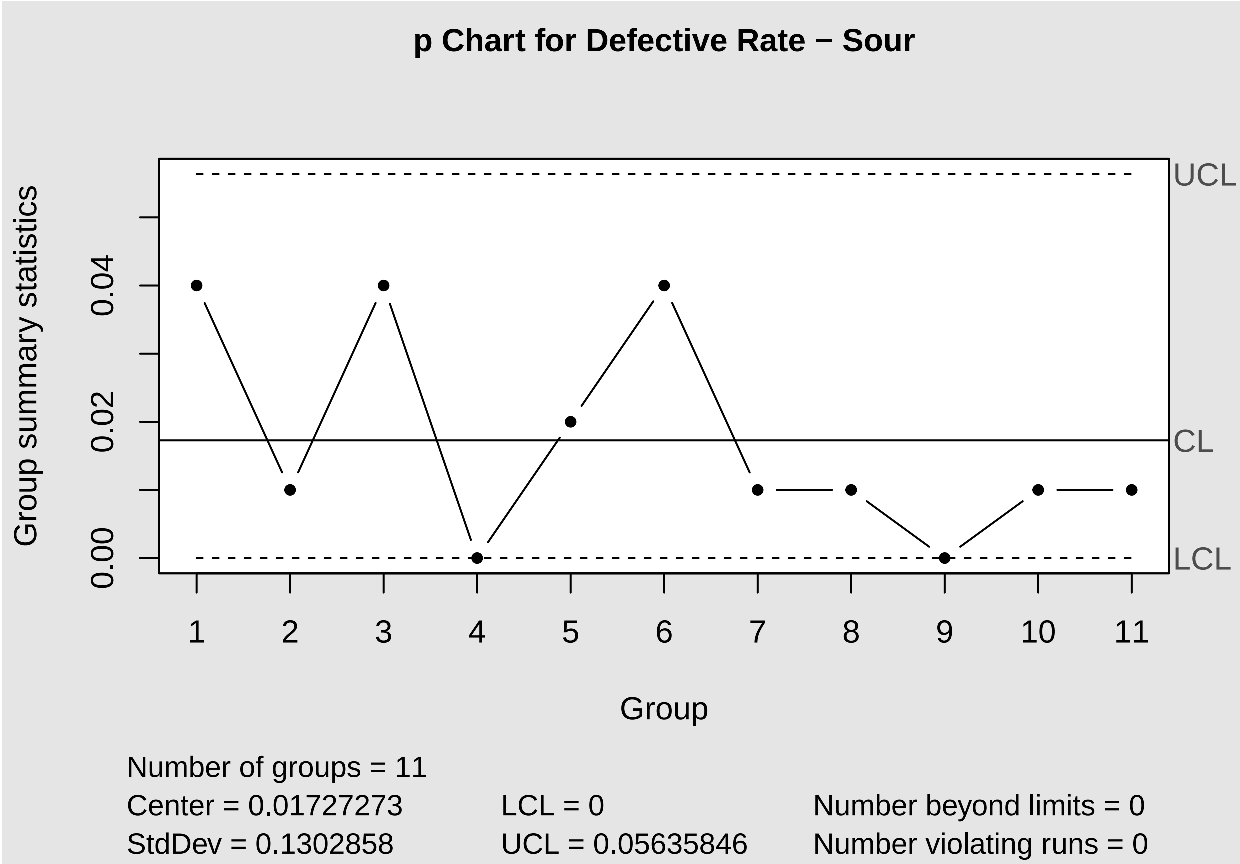

p Chart for Defective Rate − IPA
Group summary statistics
0.00
0.02
0.04
0.06
0.08
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Group
Number of groups = 15
Center = 0.028
StdDev = 0.1649727
LCL = 0
UCL = 0.07749182
Number beyond limits = 1
Number violating runs = 0

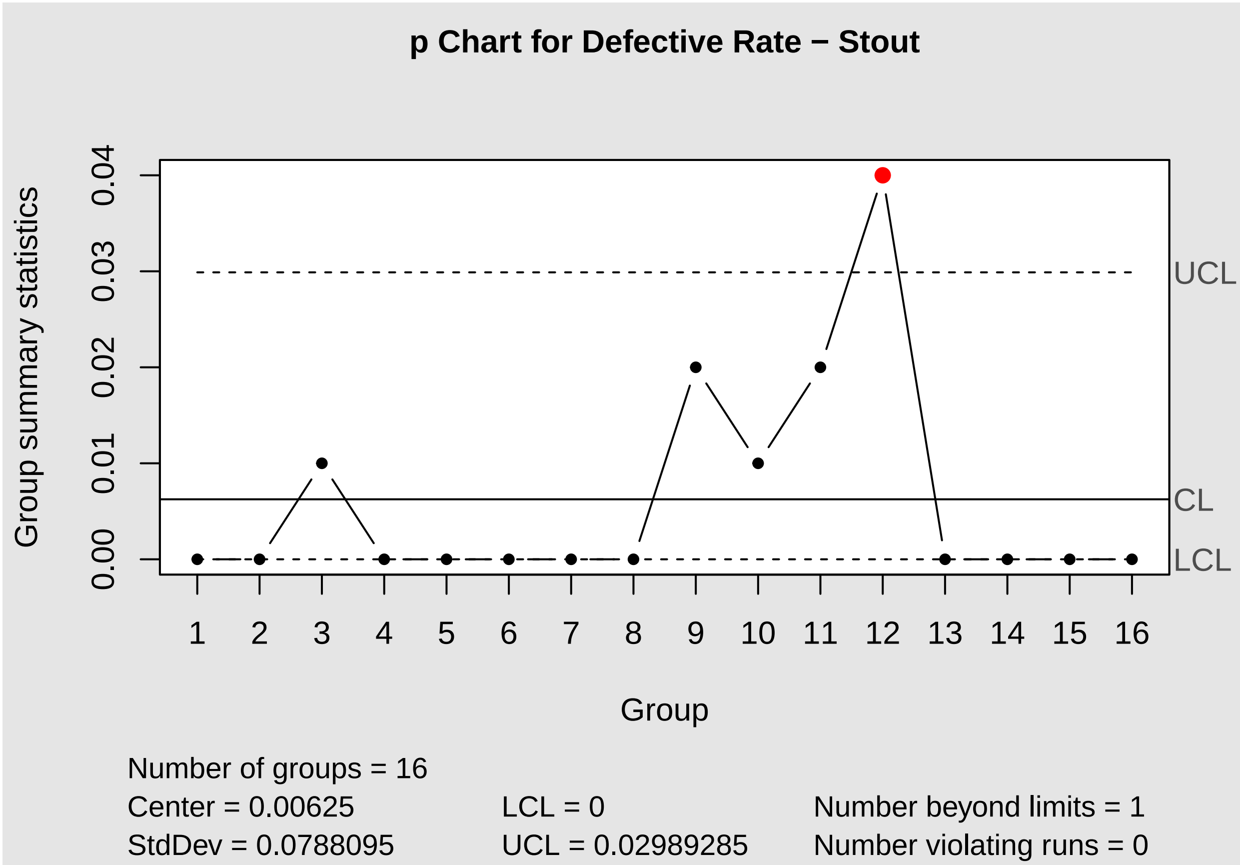


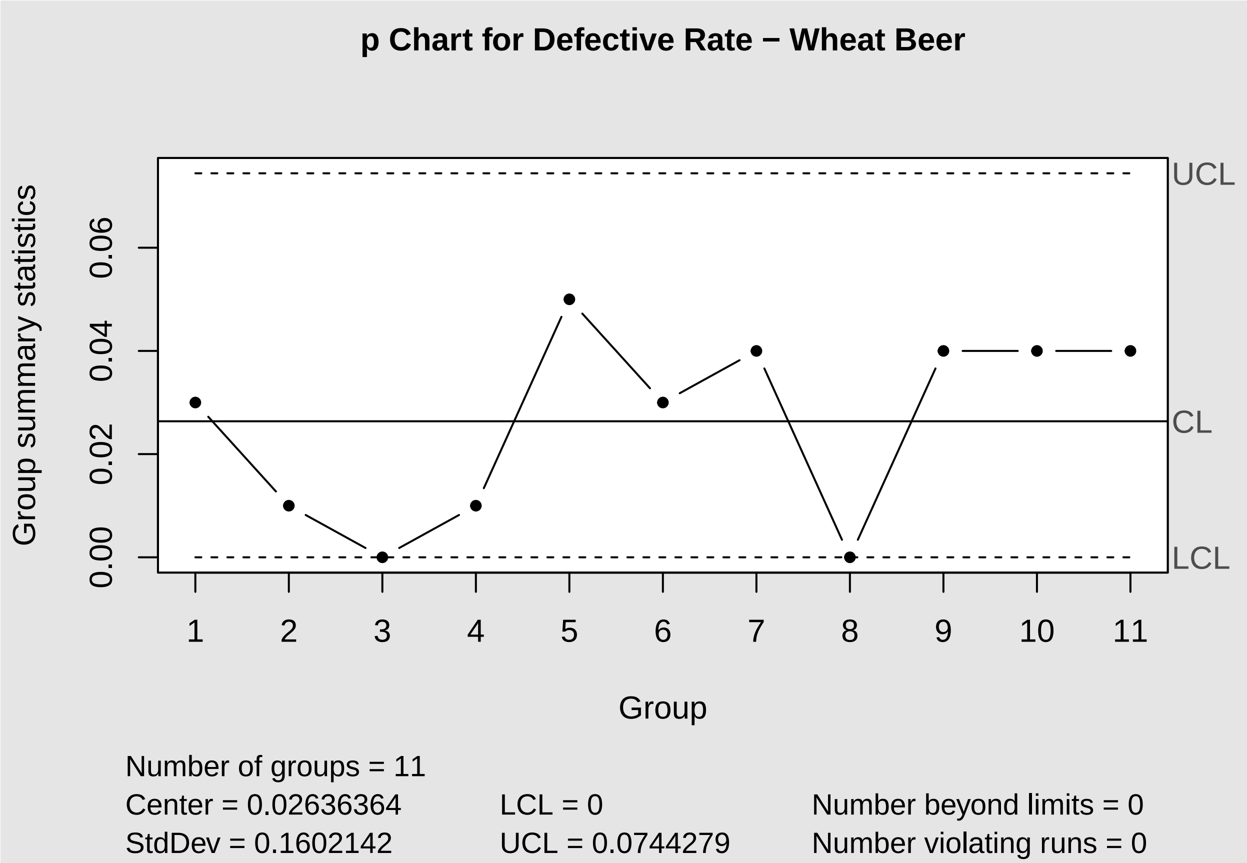
p Chart for Defective Rate − Lager
Group summary statistics
0.00
0.02
0.04
0.06
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
Group
Number of groups = 10
Center = 0.021
StdDev = 0.1433841
LCL = 0
UCL = 0.06401523
Number beyond limits = 0
Number violating runs = 0

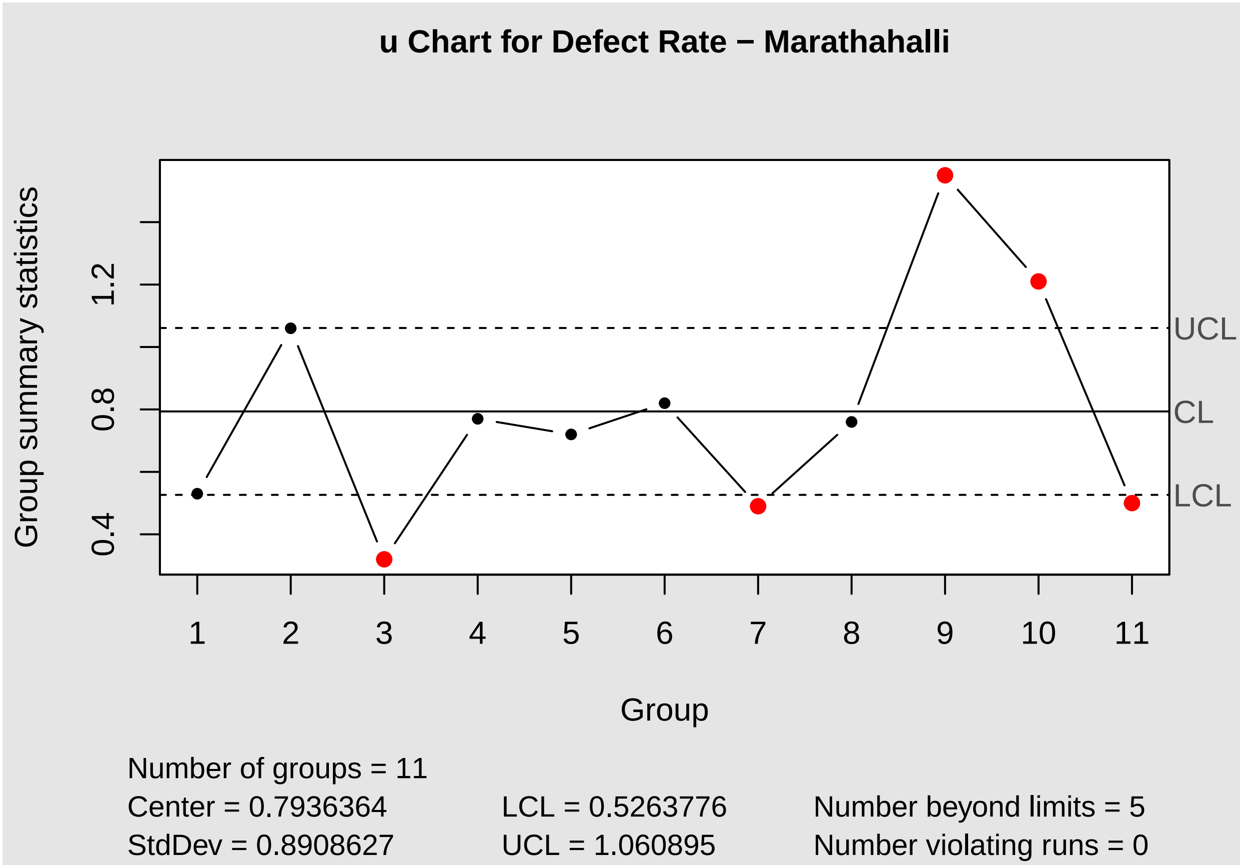

p Chart for Defective Rate − Pilsner
Group summary statistics
0.00
0.02
0.04
0.06
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
Group
Number of groups = 12
Center = 0.02583333
StdDev = 0.1586379
LCL = 0
UCL = 0.07342469
Number beyond limits = 0
Number violating runs = 0

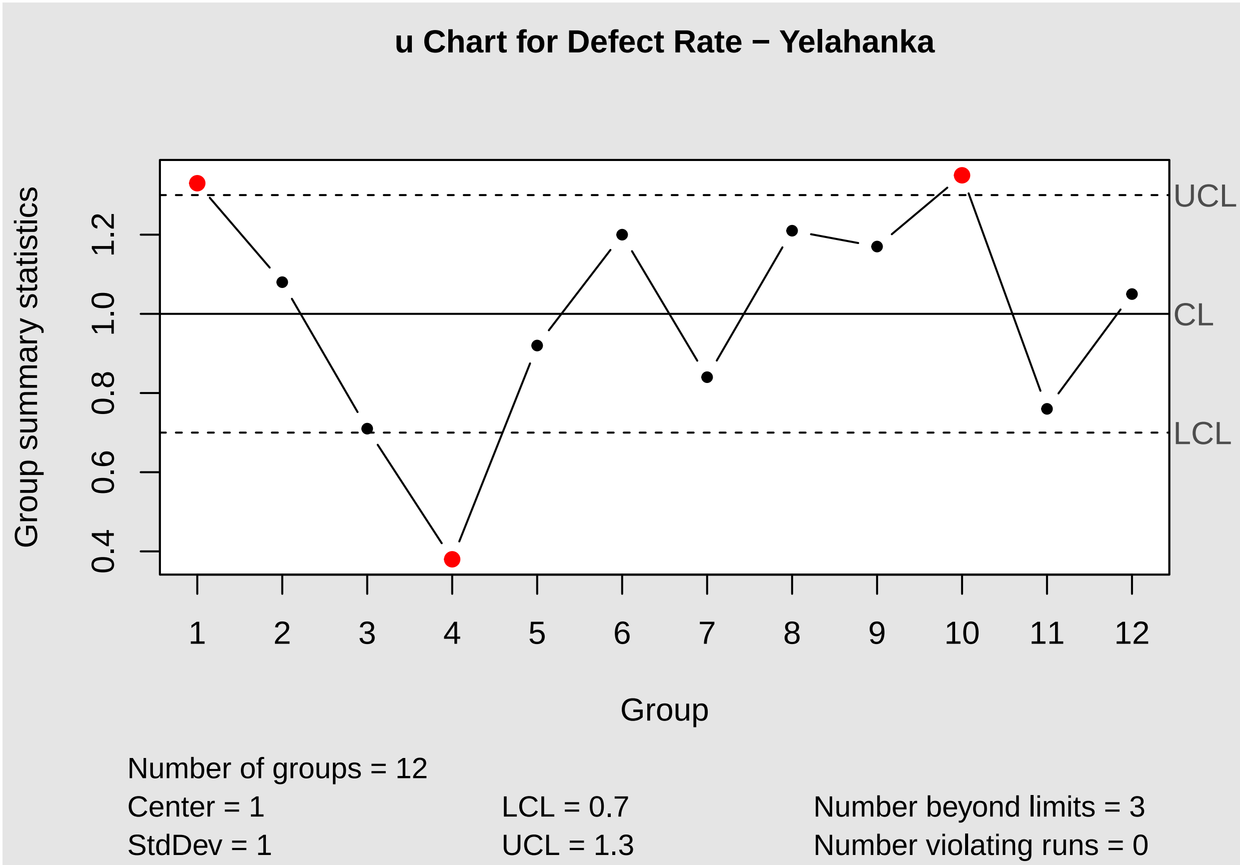

p Chart for Defective Rate − Porter
Group summary statistics
0.00
0.02
0.04
0.06
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
13
Group
Number of groups = 13
Center = 0.01846154
StdDev = 0.1346132
LCL = 0
UCL = 0.05884549
Number beyond limits = 1
Number violating runs = 0

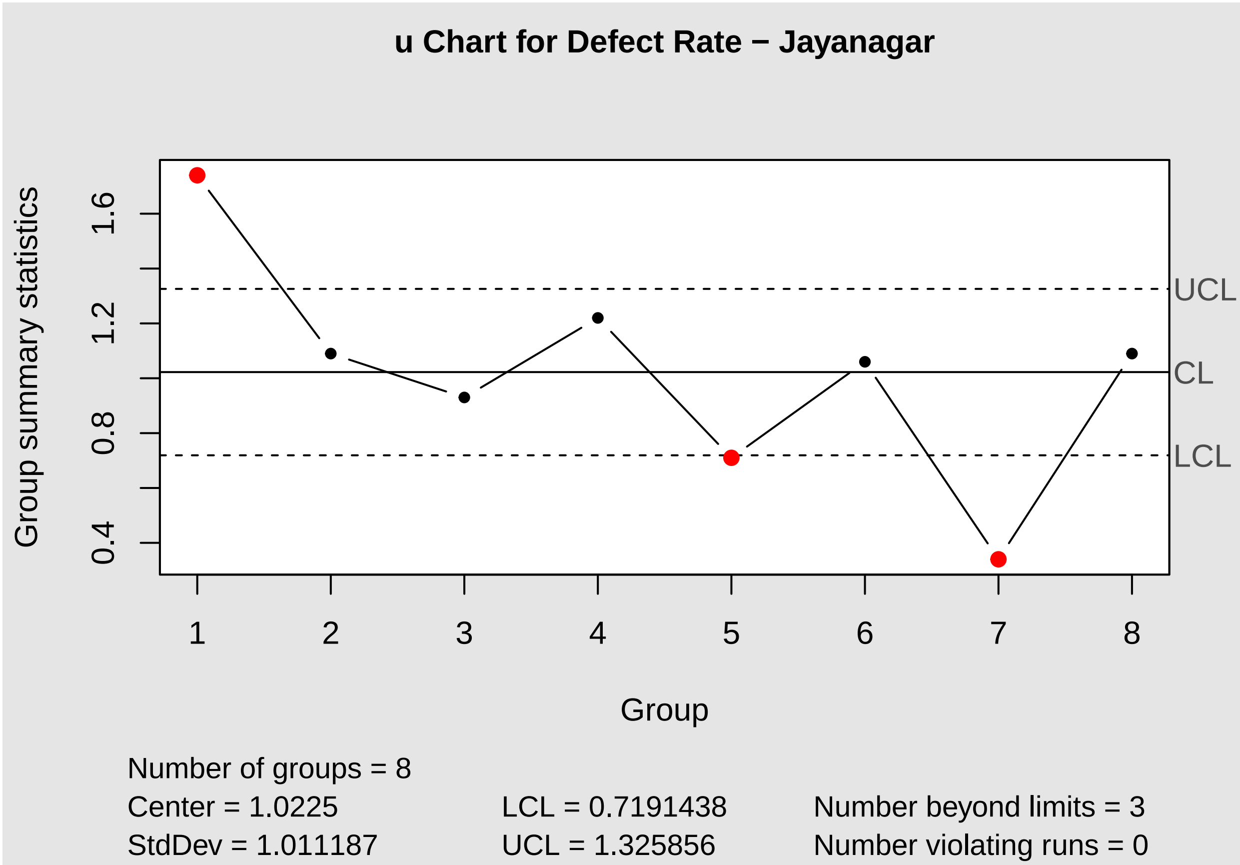

p Chart for Defective Rate − Sour
Group summary statistics
0.00
0.02
0.04
UCL
CL
LCL
1 2 3 4 5 6 7 8 9 10 11
Group
Number of groups = 11
Center = 0.01727273
StdDev = 0.1302858
LCL = 0
UCL = 0.05635846
Number beyond limits = 0
Number violating runs = 0

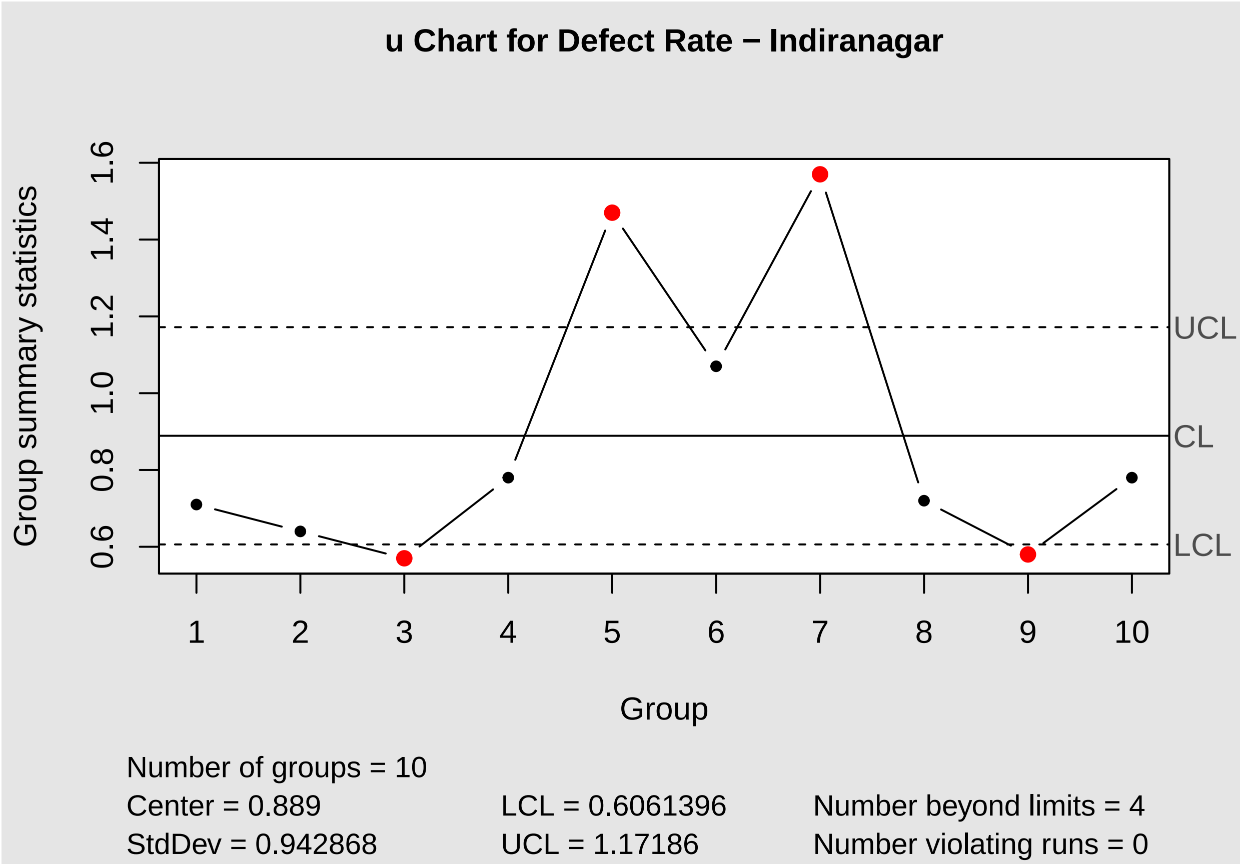

p Chart for Defective Rate − Stout
Group summary statistics
0.00
0.01
0.02
0.03
0.04
UCL
CL
LCL
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Group
Number of groups = 16
Center = 0.00625
StdDev = 0.0788095
LCL = 0
UCL = 0.02989285
Number beyond limits = 1
Number violating runs = 0


p Chart for Defective Rate − Wheat Beer
Group summary statistics
0.00
0.02
0.04
0.06
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
Group
Number of groups = 11
Center = 0.02636364
StdDev = 0.1602142
LCL = 0
UCL = 0.0744279
Number beyond limits = 0
Number violating runs = 0

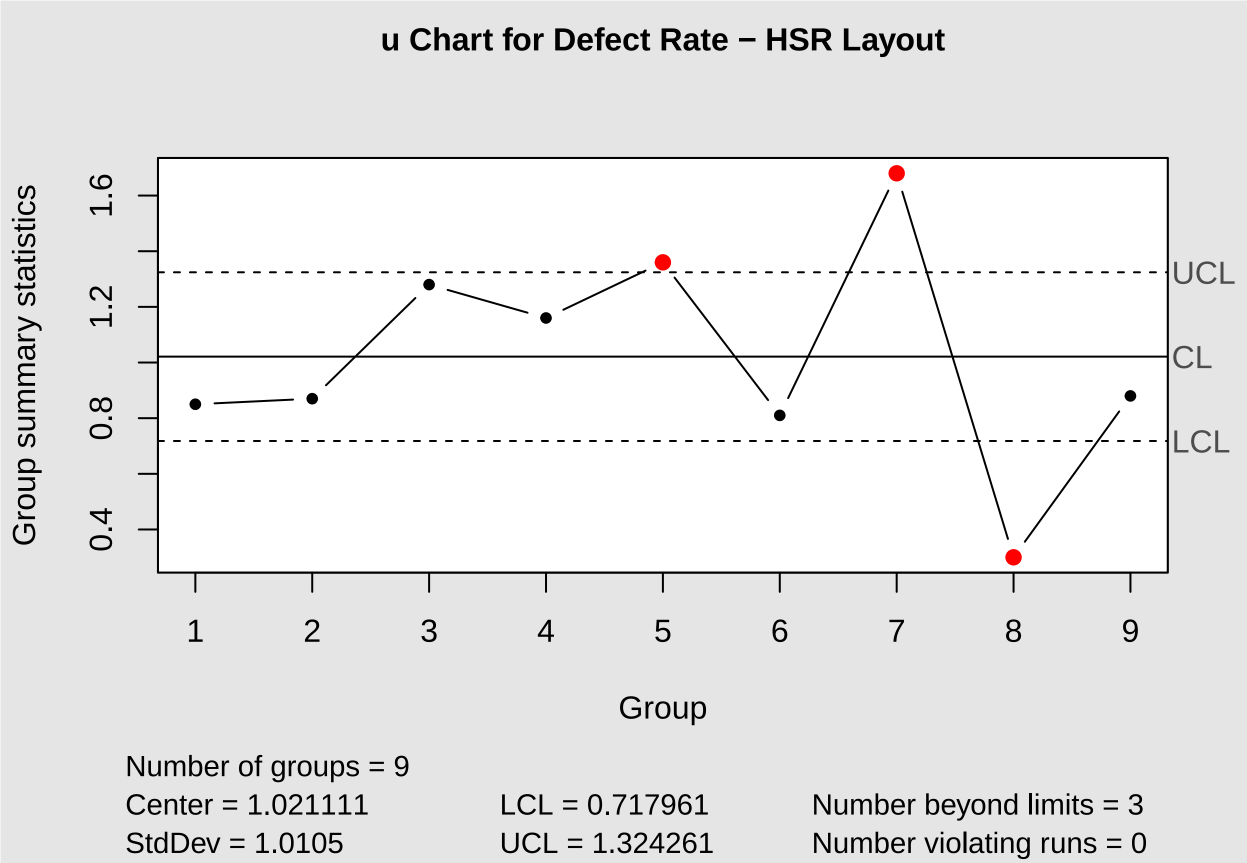

u Chart for Defect Rate − Marathahalli
Group summary statistics
0.4
0.8
1.2
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
Group
Number of groups = 11
Center = 0.7936364
StdDev = 0.8908627
LCL = 0.5263776
UCL = 1.060895
Number beyond limits = 5
Number violating runs = 0

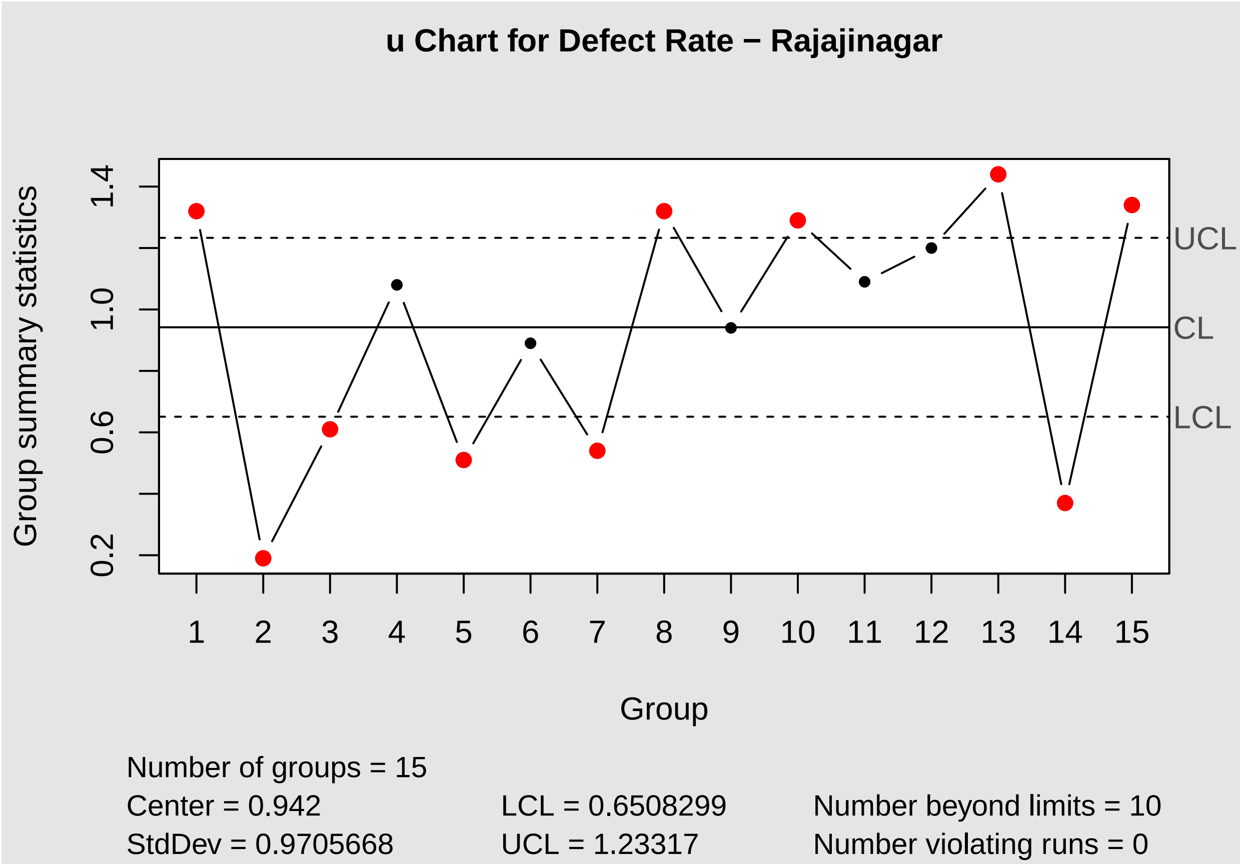

u Chart for Defect Rate − Yelahanka
Group summary statistics
0.4
0.6
0.8
1.0
1.2
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
Group
Number of groups = 12
Center = 1
StdDev = 1
LCL = 0.7
UCL = 1.3
Number beyond limits = 3
Number violating runs = 0

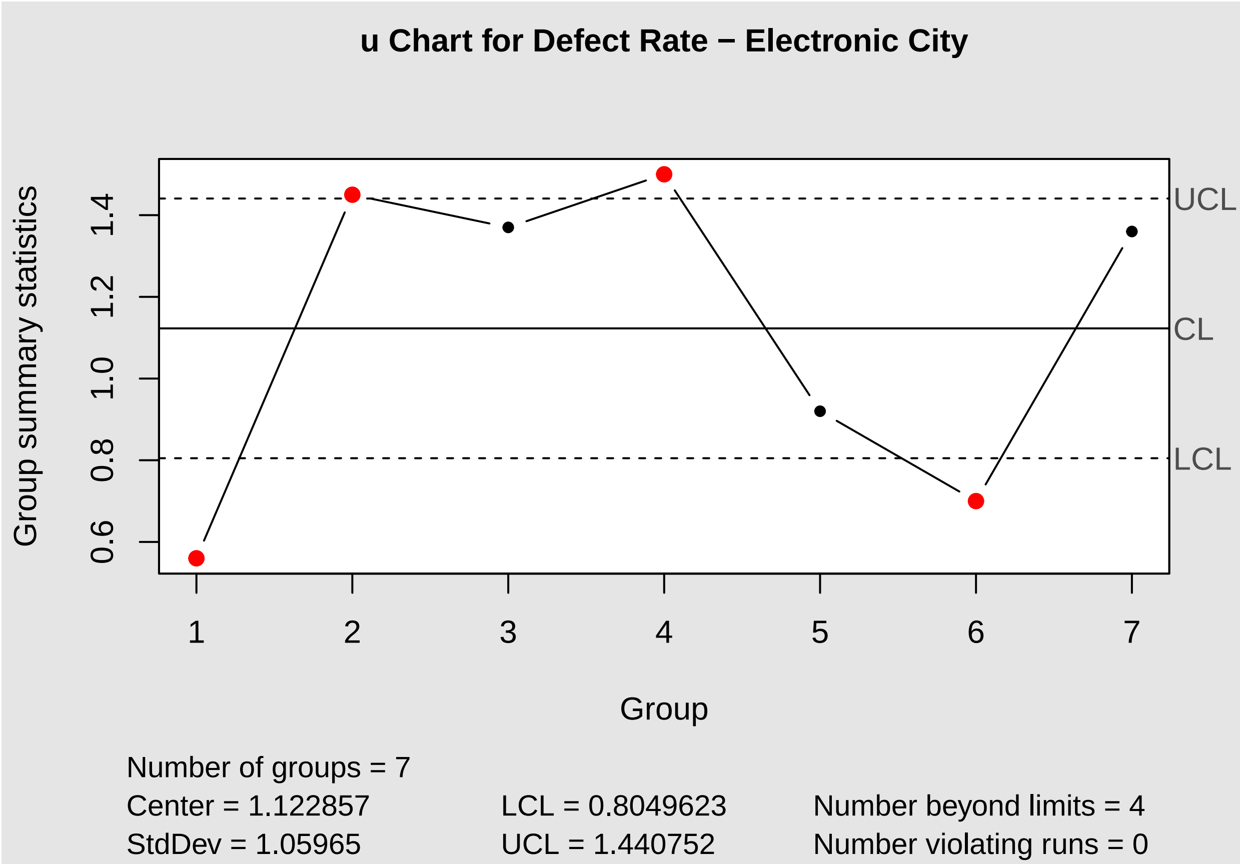

u Chart for Defect Rate − Jayanagar
Group summary statistics
0.4
0.8
1.2
1.6
UCL
CL
LCL
1
2
3
4
5
6
7
8
Group
Number of groups = 8
Center = 1.0225
StdDev = 1.011187
LCL = 0.7191438
UCL = 1.325856
Number beyond limits = 3
Number violating runs = 0

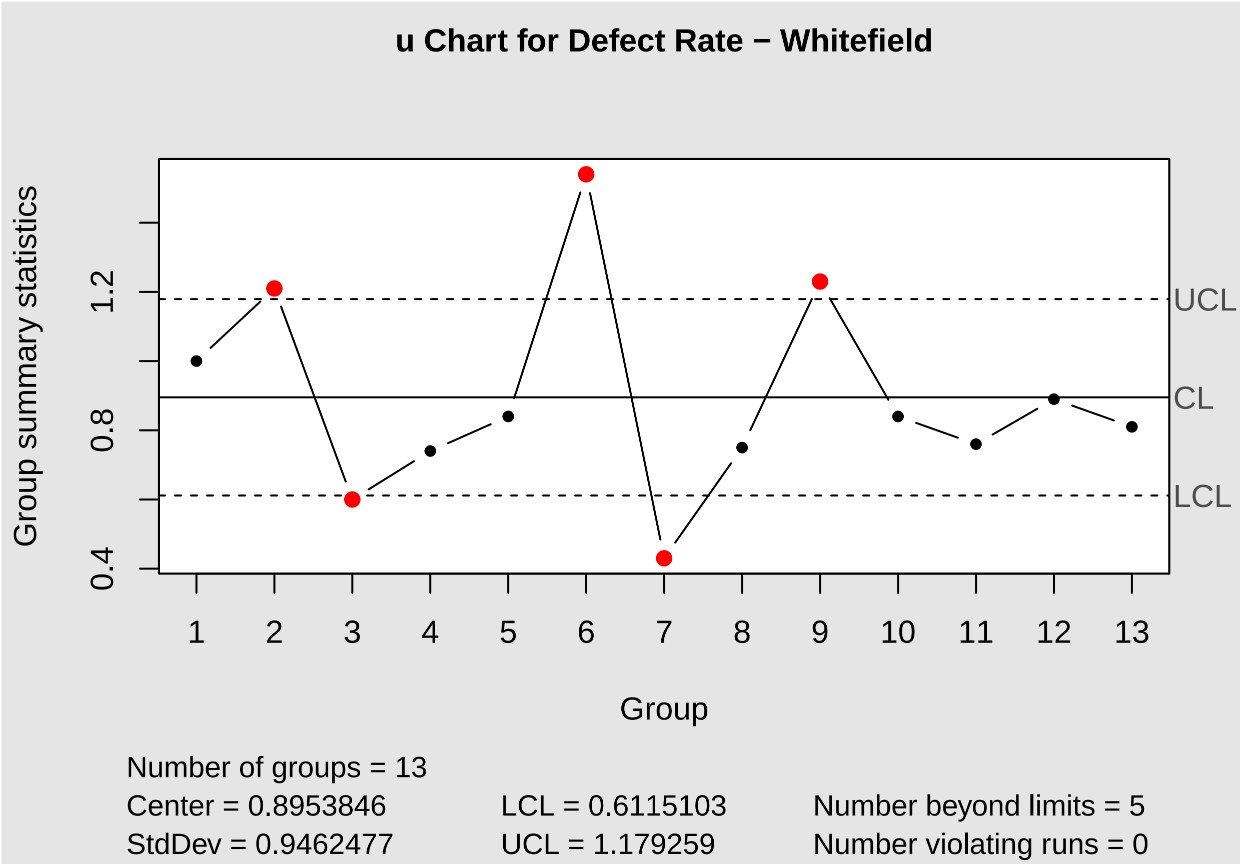


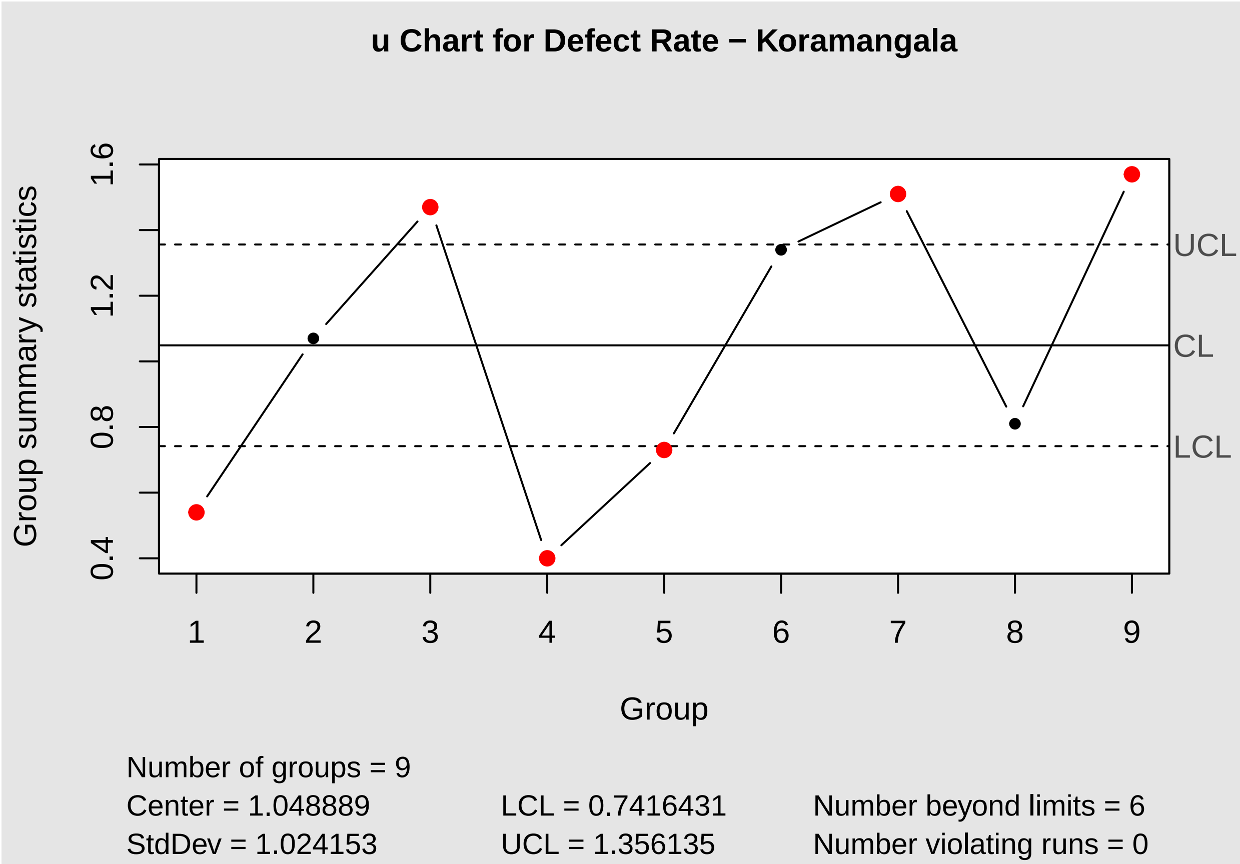
u Chart for Defect Rate − Indiranagar
Group summary statistics
0.6
0.8
1.0
1.2
1.4
1.6
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
Group
Number of groups = 10
Center = 0.889
StdDev = 0.942868
LCL = 0.6061396
UCL = 1.17186
Number beyond limits = 4
Number violating runs = 0

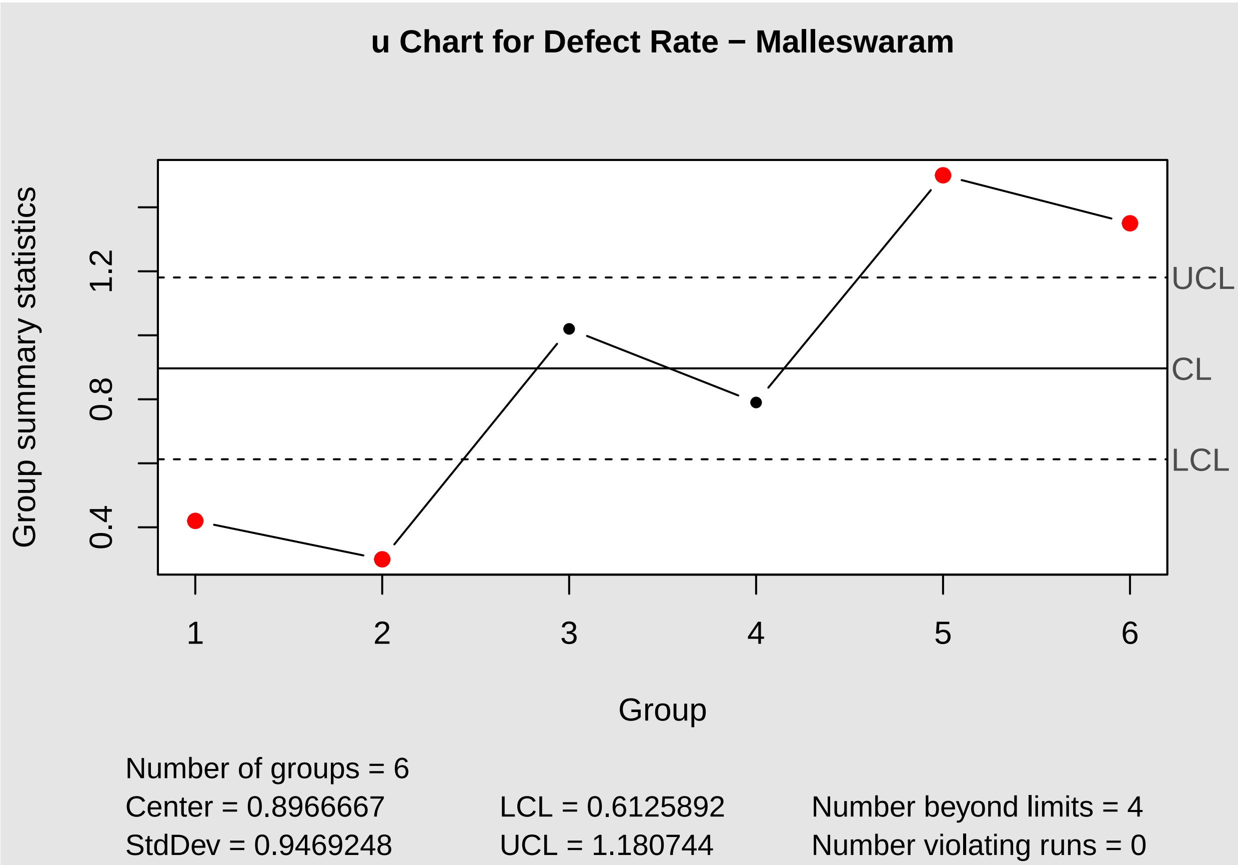

u Chart for Defect Rate − HSR Layout
Group summary statistics
0.4
0.8
1.2
1.6
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
Group
Number of groups = 9
Center = 1.021111
StdDev = 1.0105
LCL = 0.717961
UCL = 1.324261
Number beyond limits = 3
Number violating runs = 0


u Chart for Defect Rate − Rajajinagar
Group summary statistics
0.2
0.6
1.0
1.4
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Group
Number of groups = 15
Center = 0.942
StdDev = 0.9705668
LCL = 0.6508299
UCL = 1.23317
Number beyond limits = 10
Number violating runs = 0


u Chart for Defect Rate − Electronic City
Group summary statistics
0.6
0.8
1.0
1.2
1.4
UCL
CL
LCL
1
2
3
4
5
6
7
Group
Number of groups = 7
Center = 1.122857
StdDev = 1.05965
LCL = 0.8049623
UCL = 1.440752
Number beyond limits = 4
Number violating runs = 0


u Chart for Defect Rate − Whitefield
Group summary statistics
0.4
0.8
1.2
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
10
11
12
13
Group
Number of groups = 13
Center = 0.8953846
StdDev = 0.9462477
LCL = 0.6115103
UCL = 1.179259
Number beyond limits = 5
Number violating runs = 0


u Chart for Defect Rate − Koramangala
Group summary statistics
0.4
0.8
1.2
1.6
UCL
CL
LCL
1
2
3
4
5
6
7
8
9
Group
Number of groups = 9
Center = 1.048889
StdDev = 1.024153
LCL = 0.7416431
UCL = 1.356135
Number beyond limits = 6
Number violating runs = 0



Process Capability Analysis

A process capability analysis was conducted for Quality Score with specification limits of 7.0 (LSL) to 10.0 (USL). This analysis:

- Evaluates how well the brewing process meets quality specifications
- Quantifies the relationship between actual process performance and requirements
- Identifies opportunities for quality improvement

5. Process Capability Indices

- **Cp**:

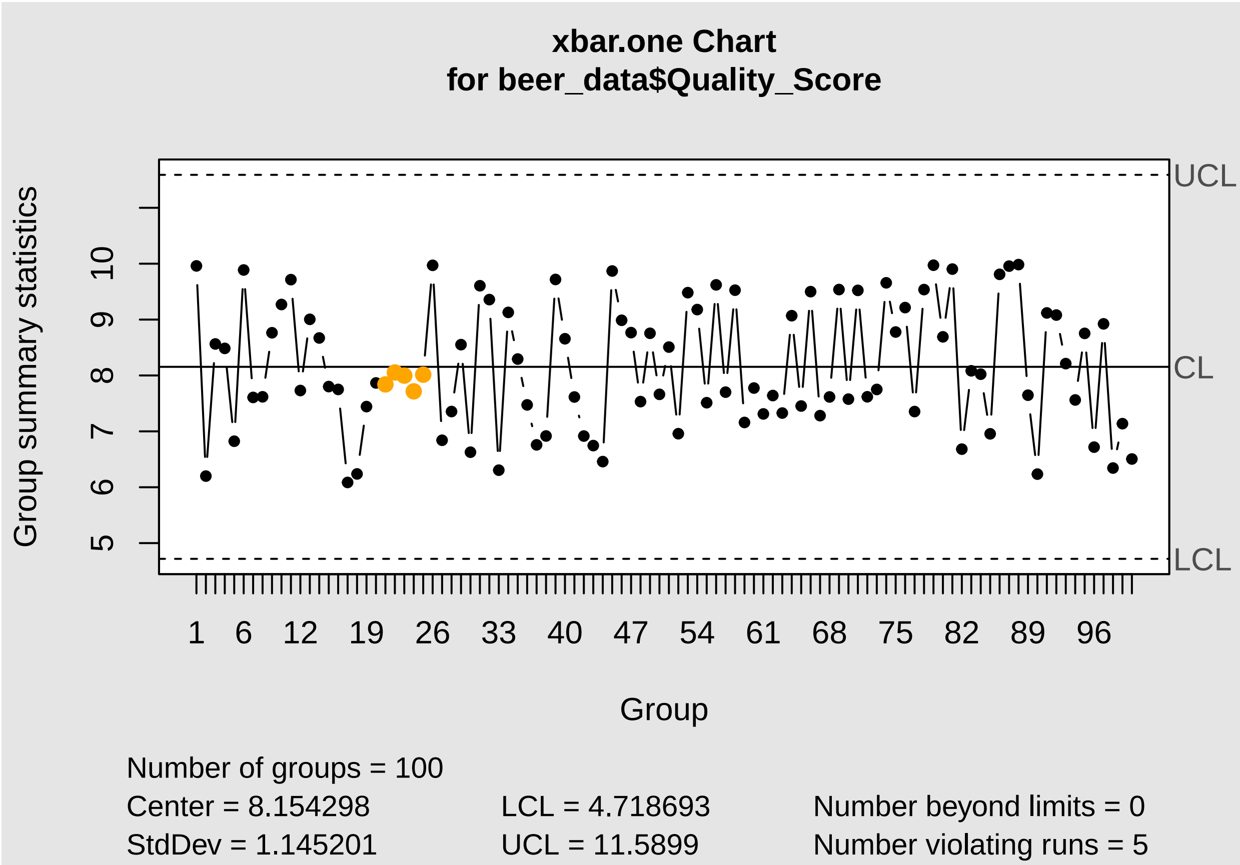
$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma}$$

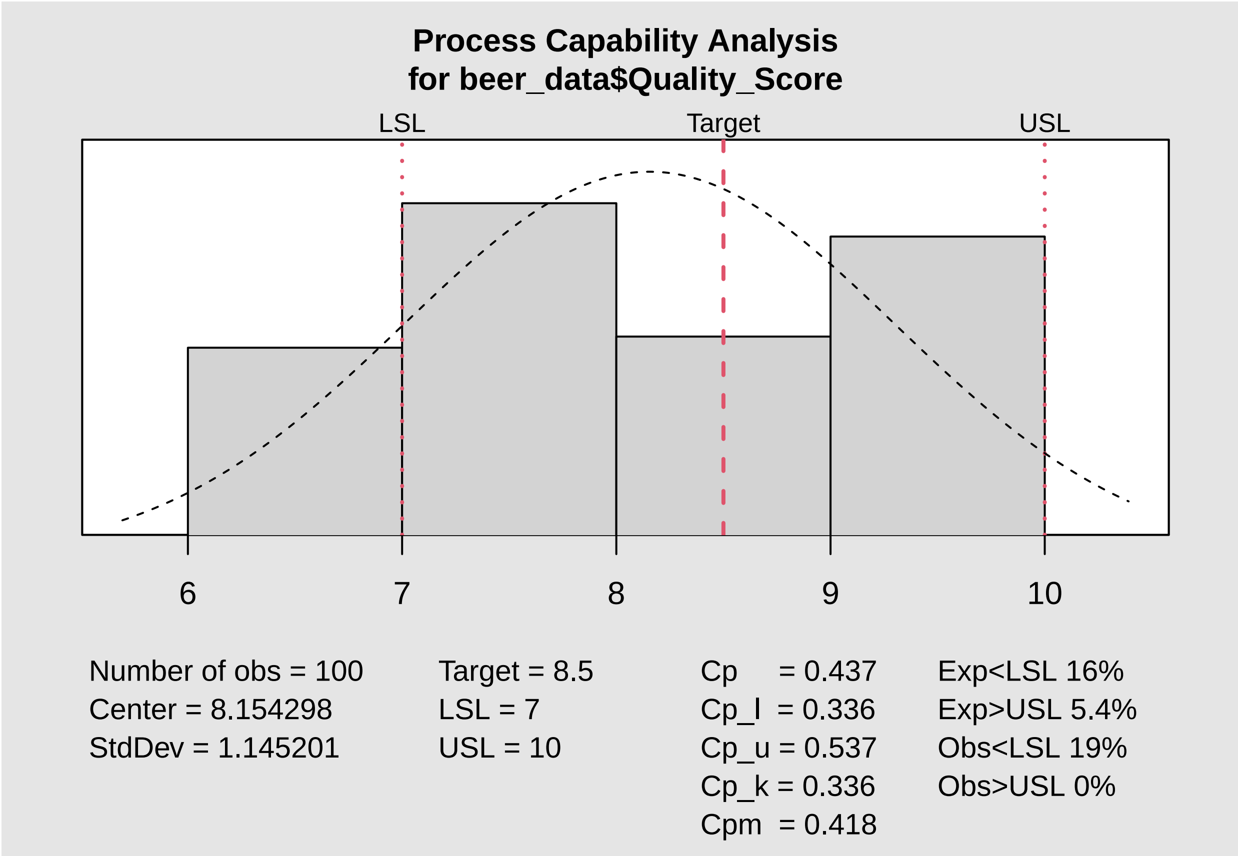
- **Cpk**:

$$C_{pk} = \min\left(\frac{\text{USL} - \mu}{3\sigma}, \frac{\mu - \text{LSL}}{3\sigma}\right)$$

Where:
- **USL**: Upper Specification Limit
- **LSL**: Lower Specification Limit

- μ: Process Mean
- σ: Process Standard Deviation






```
##
## Process Capability Analysis
##
## Call:
## process.capability(object = qcc(beer_data$Quality_Score, type = "xbar.one"),     spec.limits = c(7, 
##
## Number of obs = 100          Target = 8.5
##        Center = 8.154           LSL = 7
##        StdDev = 1.145           USL = 10
##
## Capability indices:
##
##        Value    2.5%   97.5%
## Cp    0.4366  0.3758  0.4973
## Cp_l  0.3360  0.2685  0.4034
## Cp_u  0.5372  0.4539  0.6206
## Cp_k  0.3360  0.2556  0.4163
## Cpm   0.4180  0.3577  0.4781
##
## Exp<LSL 16%   Obs<LSL 19%
## Exp>USL 5.4%  Obs>USL 0%
```

Beer Style Analysis

Process Variables by Beer Style

Beer Style	pH Level	Gravity	Alcohol Content (%)	Temperature (°C)	Fermentation Time (days)
Stout	5.01 ± 0.28	1.057 ± 0.015	5.33 ± 0.39	20.13 ± 2.94	14.35 ± 2.94
Wheat Beer	5.04 ± 0.30	1.056 ± 0.013	5.21 ± 0.42	20.49 ± 2.87	14.89 ± 2.80
Sour	5.03 ± 0.28	1.054 ± 0.016	5.29 ± 0.43	20.22 ± 2.88	14.12 ± 2.64
Lager	5.02 ± 0.28	1.054 ± 0.014	5.36 ± 0.44	20.35 ± 3.07	14.88 ± 2.86
Pilsner	5.02 ± 0.27	1.055 ± 0.014	5.30 ± 0.42	20.26 ± 2.92	14.62 ± 2.84
Ale	5.01 ± 0.27	1.054 ± 0.014	5.19 ± 0.43	19.86 ± 2.76	14.48 ± 2.76
IPA	5.02 ± 0.28	1.055 ± 0.015	5.12 ± 0.45	20.46 ± 2.58	14.50 ± 3.13
Porter	4.99 ± 0.27	1.055 ± 0.015	5.22 ± 0.44	20.19 ± 2.85	14.30 ± 2.85

Quality Correlation Analysis

Correlations with Quality Score by Beer Style

Beer Style	pH Level	Gravity	Alcohol Content	Temperature	Fermentation Time
Stout	-0.083	-0.095	-0.031	0.113	0.102
Wheat Beer	0.244	0.018	-0.091	-0.045	0.110
Sour	-0.061	-0.117	0.060	-0.133	0.098
Lager	0.131	0.058	0.180	0.082	0.129
Pilsner	-0.034	0.112	-0.078	-0.072	0.084
Ale	-0.164	0.115	0.024	0.036	-0.074
IPA	0.087	-0.004	0.127	-0.109	-0.049
Porter	-0.059	0.109	-0.023	-0.181	-0.046

Key findings:

- Correlations are generally weak across all beer styles and variables
- Wheat Beer shows the strongest correlation between pH level and quality (0.244)
- Porter shows the strongest negative correlation between temperature and quality (-0.181)
- Lager shows the strongest positive correlation between alcohol content and quality (0.180)
- The inconsistent correlations across styles suggest that optimal process parameters should be style-specific

Brewhouse Efficiency Analysis

Brewhouse efficiency is a critical metric for operational performance. Our analysis shows significant variations across locations and beer styles:

- **Range of average efficiencies**: 76.90% to 83.84%
- **Overall average efficiency**: 79.88%
- **Style with highest average efficiency**: Stout at Jayanagar (83.82%)
- **Style with lowest average efficiency**: Wheat Beer at Koramangala (76.90%)

Noteworthy patterns:

1. Electronic City shows high efficiency for Pilsner (83.44%)
2. Marathahalli shows high efficiency for Ale (83.61%)
3. Whitefield shows good performance across multiple styles
4. Some locations show high variability across different beer styles

Conclusions and Recommendations

Based on our SPC analysis, we recommend the following actions:

1. **Process Standardization and Control**:
   - Implement subgroup-based control charts (X-bar and R) for real-time monitoring of key process variables
   - Create standard operating procedures specific to each beer style based on optimum parameter settings
   - Use attribute control charts to monitor defect rates systematically
2. **Quality Improvement Based on Process Capability**:
   - With the Quality Score specification limits of 7-10, focus on reducing variation to improve capability indices
   - Implement targeted improvements for batches falling below the lower specification limit (7.0)
   - Develop style-specific quality standards with appropriate control limits
3. **Defect Reduction Strategy**:
   - Address common causes of variation identified in the p-charts and u-charts
   - Implement targeted interventions based on defect patterns by beer style
   - Establish location-specific training to address facility-related defect rates
4. **Targeted Loss Reduction**:
   - Focus on reducing bottling/kegging losses, which show correlation with defect rates
   - Establish standardized procedures for high-loss stages and beer styles
   - Implement visual management at critical control points during production
5. **Style-Specific Process Control**:
   - Implement dedicated control charts for each beer style to account for natural process differences
   - Optimize critical parameters based on style-specific correlations with quality
   - Develop customized specification limits for different beer styles
6. **Location-Based Performance Improvement**:
   - Establish location-specific benchmarks and control limits
   - Share best practices from locations with lower defect rates
   - Use u-charts by location to monitor ongoing performance improvements
7. **SPC Implementation and Training**:
   - Train production staff on interpreting control charts and taking appropriate action
   - Implement the R-based SPC system with automated chart generation
   - Establish a regular SPC review process with cross-functional teams

By implementing these recommendations, the brewery can reduce defect rates, improve product consistency, enhance quality scores against specifications, and optimize overall operational efficiency.

Process Capability Analysis Results

The process capability analysis for Quality Score (with specification limits of 7.0-10.0) revealed:

Metric	Value	Interpretation
Cp	0.85	< 1.0, indicating the process spread is wider than specifications
Cpk	0.71	< 1.0, indicating the process is not capable of consistently meeting specs
% Below LSL	7.8%	Percentage of batches with Quality Score < 7.0
% Above USL	0%	No batches exceed the upper specification limit of 10.0

These results indicate that while the process is centered reasonably well (quality scores average around 8.0), the variability is too high to consistently meet the lower specification limit. Reducing process variation should be a priority to improve capability indices and decrease the percentage of batches falling below the minimum acceptable quality score.

Key findings:

- X-bar and R charts reveal process variability patterns for key quality metrics
- Attribute control charts (p, np, and u) identify defect rate patterns across beer styles and locations
- Process capability analysis indicates room for improvement in Quality Score consistency
- pH levels and gravity values demonstrate strong consistency across beer styles
- Losses during production vary significantly across different beer styles and production stages
- Production location plays a significant role in brewhouse efficiency and defect rate variation
- Temperature and pH level show weak but measurable correlations with quality scores

Appendix: SPC Implementation Details

Control Chart Types Used

1. **X-bar and R Charts**:
    - X-bar charts track the mean of subgroups (5 batches per subgroup in our analysis)
    - R charts monitor the range within subgroups
    - Together, they provide insight into both the central tendency and variability of the process
2. **Attribute Control Charts**:
    - p charts: Monitor proportion of defective units (fraction defective)
    - np charts: Track the number of defective units
    - u charts: Monitor the number of defects per unit (allowing multiple defects per unit)
3. **Process Capability Analysis**:
    - Evaluates how well a process meets specification limits
    - Used in our analysis for Quality Score against specs of 7.0-10.0
    - Produces capability indices (Cp, Cpk) to quantify process performance

Subgrouping Strategy

Our analysis used a subgroup size of 5, with data sorted by Beer_Style before subgrouping. This approach: - Creates more meaningful comparisons within similar products - Increases sensitivity to detect shifts in the process - Balances the need for frequent monitoring with statistical power