经过分析，此数据集的以下列为标称属性：

'Permit Type', 'Block', 'Lot', 'Street Number', 'Street Number Suffix', 'Street Name', 'Street Suffix', 'Current Status', 'Structural Notification', 'Voluntary Soft-Story Retrofit', 'Fire Only Permit', 'Existing Use', 'Proposed Use', 'Plansets', 'TIDF Compliance', 'Existing Construction Type','Proposed Construction Type', 'Site Permit', 'Supervisor District', 'Neighborhoods - Analysis Boundaries'

以下列为数值属性：

'Number of Existing Stories', 'Number of Proposed Stories', 'Estimated Cost', 'Revised Cost', 'Existing Units', 'Proposed Units'

## 数据摘要：

1. 标称属性

部分统计结果如下：

标称属性 <Permit Type> 频数统计

|  |  |
| --- | --- |
| value | count |
| 8 | 178844 |
| 3 | 14663 |
| 4 | 2892 |
| 2 | 950 |
| 6 | 600 |
| 7 | 511 |
| 1 | 349 |
| 5 | 91 |

其余数据见文件name\_cat\_out\_bp.txt

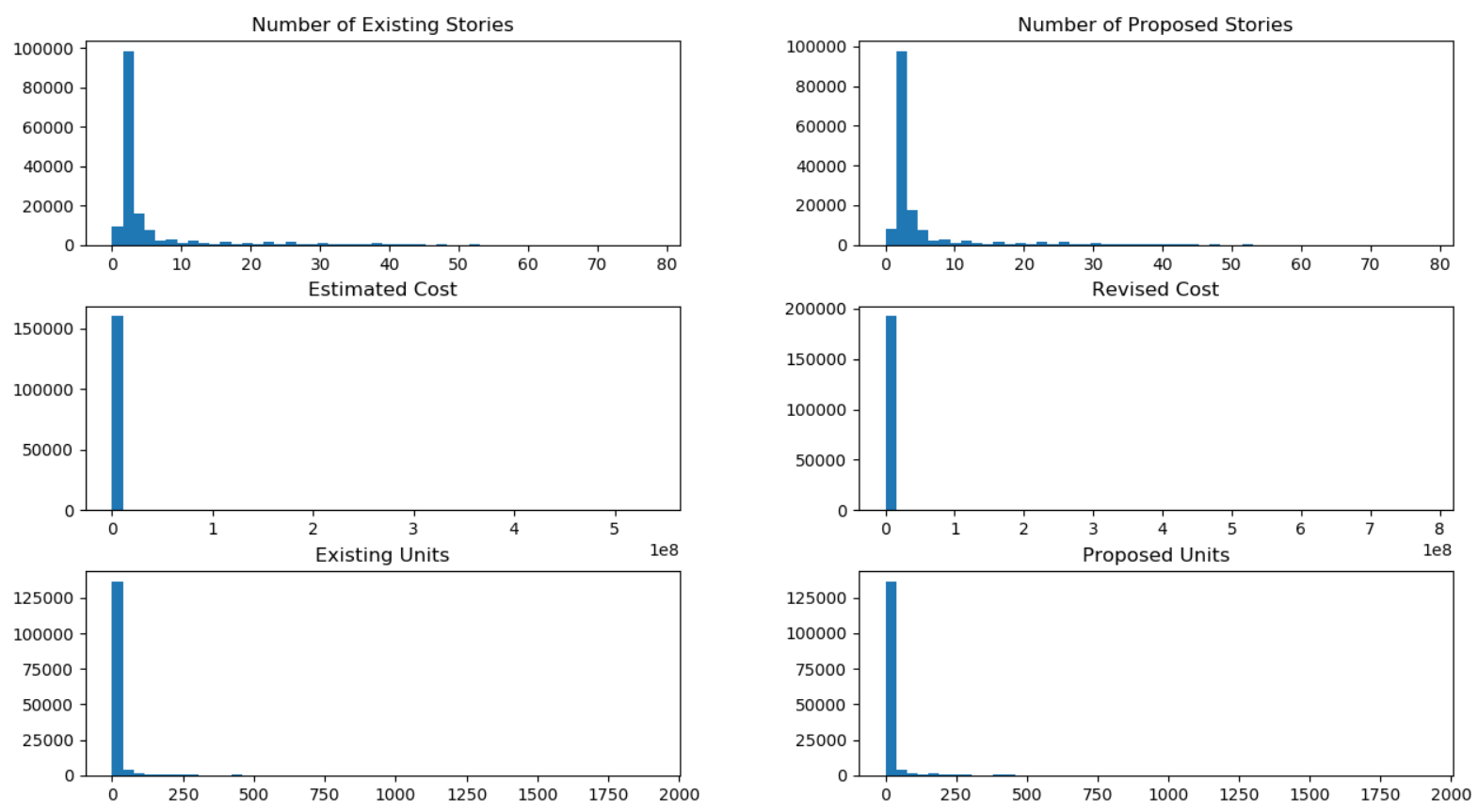
1. 数值属性

见下表

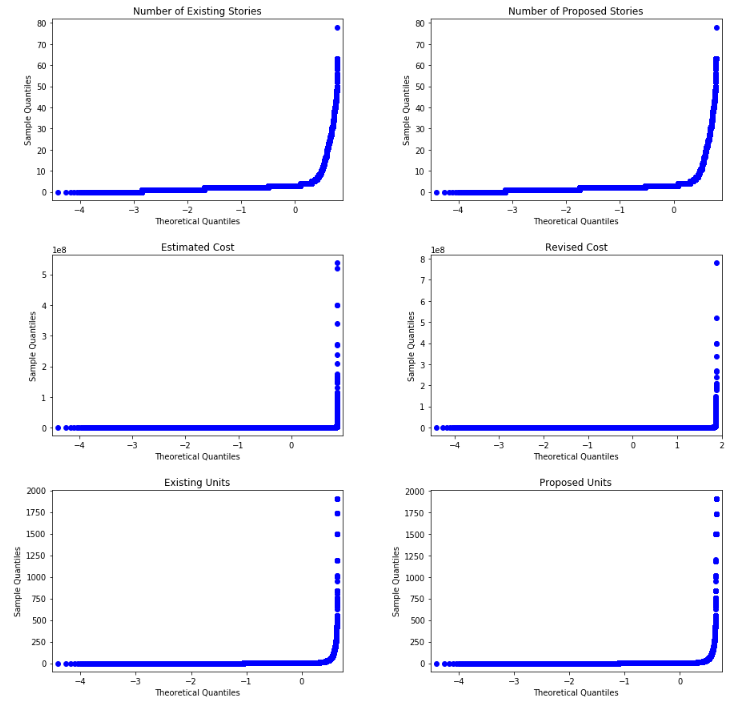
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | max | min | mean | 50% | 25% | 75% | NaN |
| Number of Existing Stories | 78 | 0 | 5.705773 | 3 | 2 | 4 | 42784 |
| Number of Proposed Stories | 78 | 0 | 5.745043 | 3 | 2 | 4 | 42868 |
| Estimated Cost | 537958646 | 1 | 168955.4 | 11000 | 3300 | 35000 | 38066 |
| Revised Cost | 780500000 | 0 | 132856.2 | 7000 | 1 | 28707.5 | 6066 |
| Existing Units | 1907 | 0 | 15.66616 | 1 | 1 | 4 | 51538 |
| Proposed Units | 1911 | 0 | 16.51095 | 2 | 1 | 4 | 50911 |

## 数据可视化

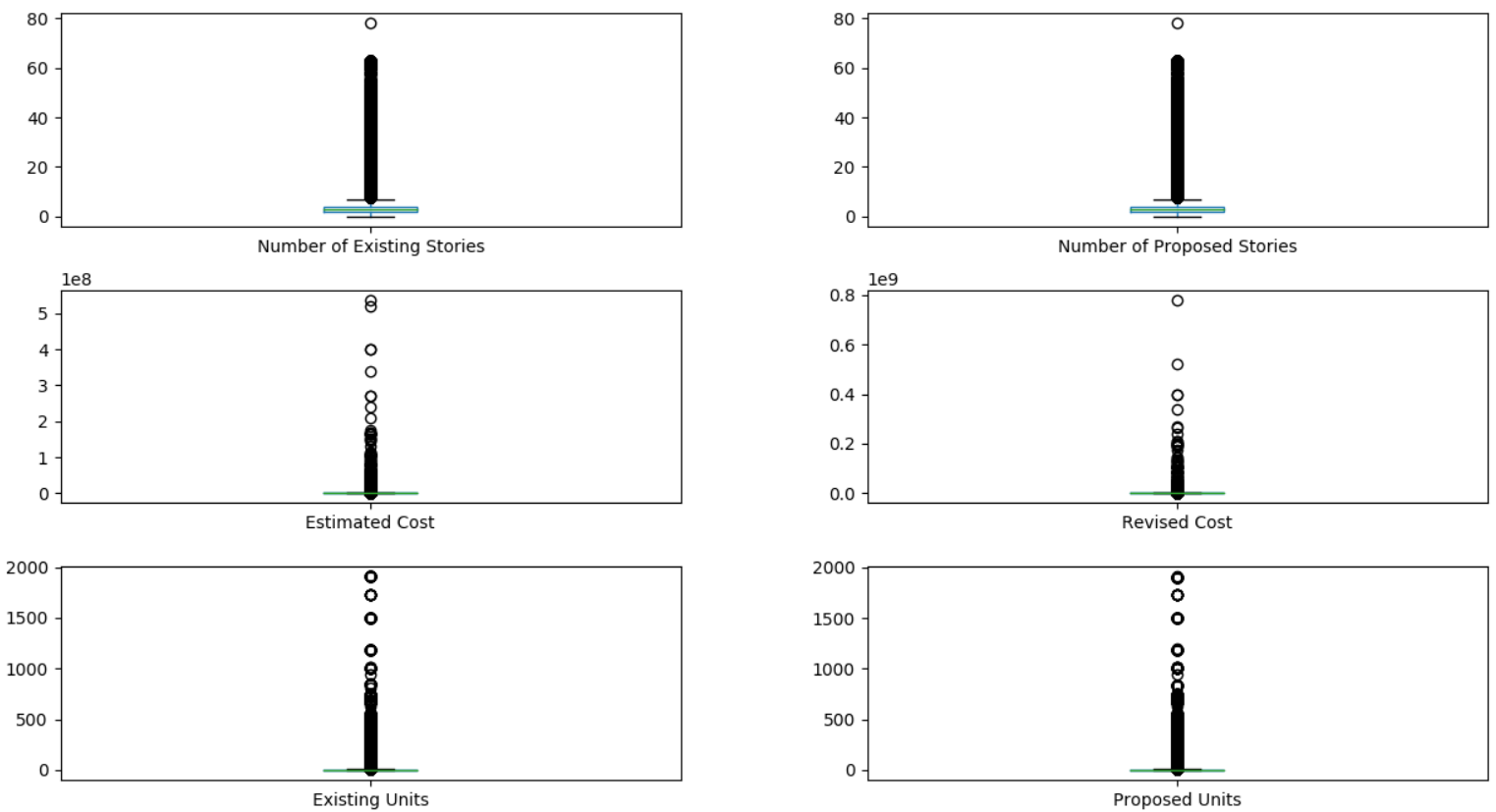
1. 直方图



1. Q-Q图



1. 盒图



## 缺失值处理

经分析，以下字段无缺失：Permit Number，Permit Type，Permit Type Definition，Permit Creation Date，Block，Lot，Street Number，Street Name，Current Status，Current Status Date，Filed Date，Record ID 。

以下字段不具备填充价值：Unit，Unit suffix，Description，Issued Date，Completed Date，First Construction Document Date，Permit Expiration Date，Existing Construction Type Description，Proposed Construction Type Description，Zipcode，Location，Street Number Suffix，Street Name Suffix，Existing Use，Existing Units，Proposed Use，Proposed Units，Plansets，Existing Construction Type，Proposed Construction Type，Supervisor District，Neighborhoods - Analysis Boundaries，Number of Existing Stories，Number of Proposed Stories，Estimated Cost，Revised Cost。上述字段为建筑的门牌号、日期等描述类字段，没有填充意义。

可填充的标称属性字段：Structural Notification，Voluntary Soft-Story Retrofit，Fire Only Permit，TIDF Compliance，Site Permit。这些字段除TIDF Compliance外均为布尔型，Y表示是，空表示否，因此可以用N填充。TIDF Compliance字段只有两条记录不为空，一条值是Y，一条是P，或许空也表示否，可用N填充。

填充后结果如下：

Structural Notification

|  |  |
| --- | --- |
| value | count |
| N | 191978 |
| Y | 6922 |

Voluntary Soft-Story Retrofit

|  |  |
| --- | --- |
| value | count |
| N | 198865 |
| Y | 35 |

Fire Only Permit

|  |  |
| --- | --- |
| value | count |
| N | 180073 |
| Y | 18827 |

TIDF Compliance

|  |  |
| --- | --- |
| value | count |
| N | 198898 |
| Y | 1 |
| P | 1 |