

# Assignment 1

Matthias  
Bennett

January 30th, 2024

## 1 Statistical Distribution of Attributes

	<b>radius_mean</b>	<b>smoothness_mean</b>	<b>texture_se</b>
<b>mean</b>	14.221321	0.096536	1.200766
<b>median</b>	13.46000	0.09646	1.07300
<b>Q1</b>	11.760000	0.086770	0.828200
<b>Q3</b>	16.130000	0.105400	1.471000

## 2 Outlier Identification

```
upper_bound = data["radius_mean"]["Q3"] + (data["radius_mean"]["IQR"] * 1.5)
lower_bound = data["radius_mean"]["Q1"] - (data["radius_mean"]["IQR"] * 1.5)
```

[27] ✓ 0.0s Python

```
lower_bound
```

[34] ✓ 0.0s Python

... 5.205000000000001

```
upper_bound
```

[35] ✓ 0.0s Python

... 22.685

```
rad = sda.radius_mean
# filter = rad > upper_bound or rad < lower_bound
filtered_rad = rad[(rad > upper_bound) | (rad < lower_bound)]
filtered_rad
```

[32] ✓ 0.0s Python

... 82 25.22  
122 24.25  
164 23.27  
180 27.22  
202 23.29  
212 28.11  
236 23.21  
339 23.51  
352 25.73  
Name: radius\_mean, dtype: float64

```
sda.sort_values(by=["radius_mean"],ascending=False)
```

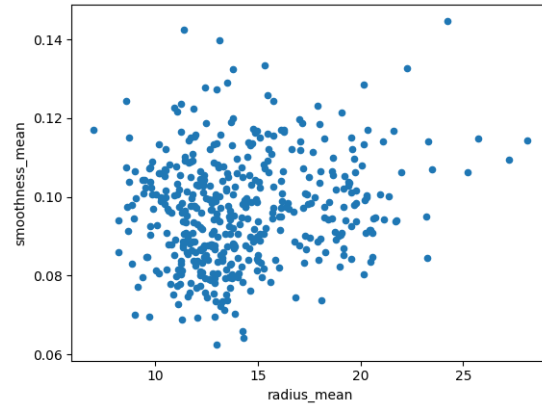
[36] ✓ 0.0s Python

	radius_mean	smoothness_mean	texture_se	symmetry_mean	compactness_mean
212	28.110	0.11420	1.4760	0.1648	0.15160
180	27.220	0.10940	1.4810	0.1800	0.19140
352	25.730	0.11490	0.8509	0.1956	0.23630
82	25.220	0.10630	1.4740	0.1829	0.26650
122	24.250	0.14470	3.1200	0.2655	0.28670
...	...	...	...	...	...
61	8.598	0.12430	2.0670	0.1828	0.08963
314	8.597	0.10740	2.7770	0.2163	0.05847
151	8.219	0.09405	1.9620	0.2222	0.13050
46	8.196	0.08600	0.9567	0.1769	0.05943
101	6.981	0.11700	1.5080	0.1930	0.07568

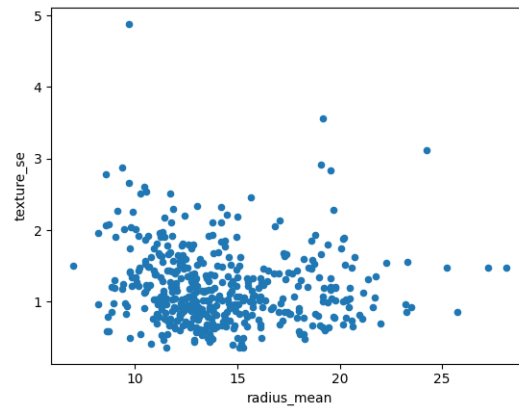
449 rows × 5 columns

### 3 Correlations

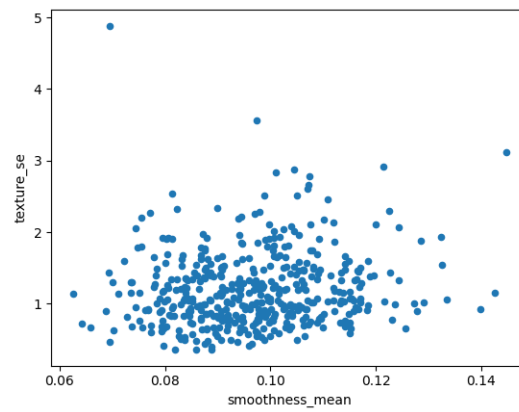
In this section we will explore the correlations between different attributes



No Correlation

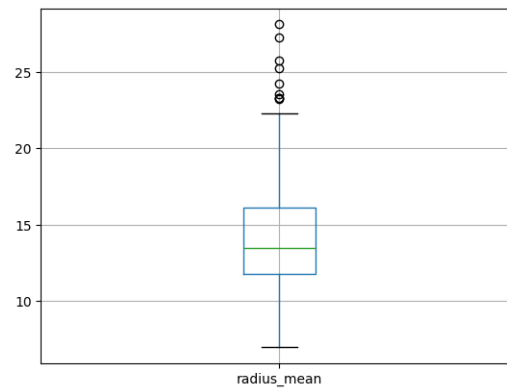


No Correlation

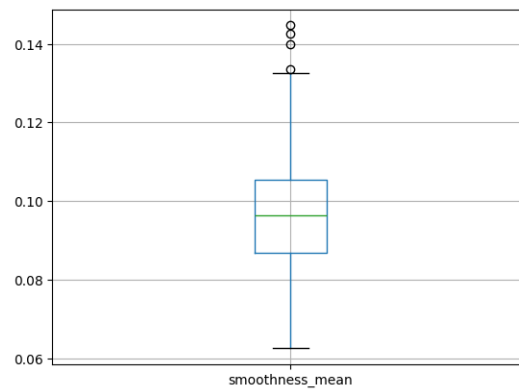


No Correlation

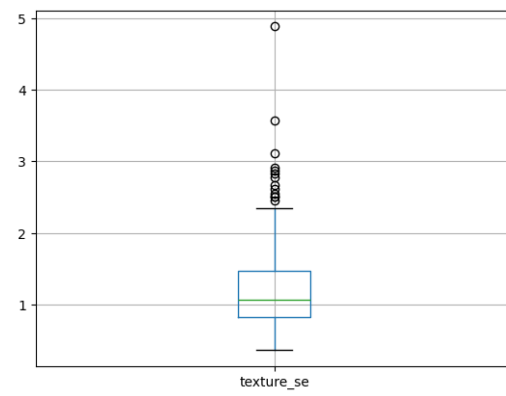
## 4 Boxplots



Description -Something about all the distributions having outliers above the Q3, all follow this pattern - ask about



Description



Description