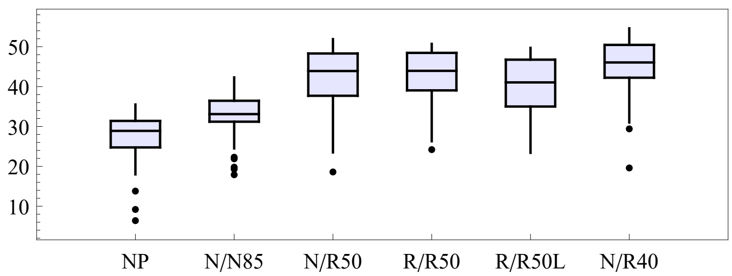
Name: Rufus Petrie

Final Problem 6: Lifespan Study

Part (a) - Summary Information

```
BoxPlot[lifespans[[1]], lifespans[[2]], lifespans[[3]],
 lifespans[[4]], lifespans[[5]], lifespans[[6]],
 Labels → labelList,
 AspectRatio → 1 / 3]
```



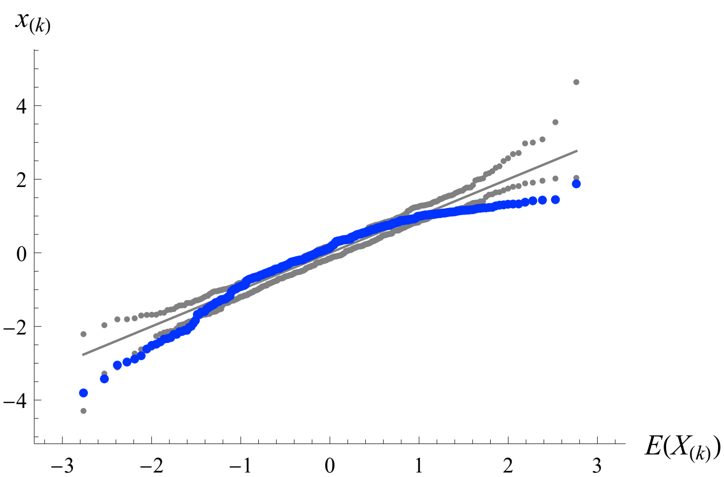

```
TableForm[Data,
 TableHeadings → {varNames, labelList}]
```

	NP	N/N85	N/R50	R/R50	R/R50L	N/R4
Sample Size	49	57	71	56	56	60
Outliers	3	5	1	1	0	2
Means	27.402	32.6912	42.2972	42.8857	39.6857	45.1
Medians	28.9	33.1	43.9	43.95	41.05	46.(
SDs	6.1337	5.1253	7.76819	6.68315	6.99169	6.7(

A few things jump out from the summary information. Firstly, it appears that caloric limits corresponded with longer lifespans in the data. Additionally, all of the means for the different groups seem somewhat close to their outliers. However, I am concerned about the lack of positive outliers. All in all, there were 12 outliers, which isn't that concerning considering that there were 349 observations (this corresponds to 3.44% of the sample), but something appears be happening with the lower results of the rats, particularly those with no restrictions. The ratio of the highest to the lowest standard deviation equals (7.76/5.12)=1.512, which indicates that there shouldn't be any issues while using parametric methods,

Part (b) - Group Differences

```
standardized = Flatten[Map[Standardize, lifespans]];
QuantileComparisonPlot[NormalDistribution[0, 1], standardized,
 SimulationBands → True]
```



Upon viewing the standardized residual plot, it appears that the normality assumption may not be so wise. For high and low predicted residuals, there are significant deviations from the prediction bands. Thus, because of this and the strange outliers, I will be using a non-parametric approach for this analysis. Because the data doesn't look like it's normally distributed, I will be using the Kruskal-Wallis test to determine if group effects have any impact on lifespan.

```
ranks = N[Ranks[lifespans]];
KruskalWallisTest[lifespans]
```

Statistic:	PValue:	Distribution:
159.013	0.	ChiSquare(df=5)

Because the Kruskal-Wallis test has a p-value of 0, we can conclude that different groups of rats had different lifespans.

Part (c) - Group Comparisons

```
RankSumTest[lifespans[[2]], lifespans[[3]],
 TwoSided → True,
 ExactDistribution → True]
```

Rank Sum Statistic:	TwoSided P Value:	Sampling Distribution:
2271.	0	Exact

Question 1 corresponds to testing for group differences between the N/N85 group and the N/R50 group because both were fed normally before weaning and restricted to 85 and 50 calorie diets after weaning, respectively. As we can see from the non-parametric rank sum test above, the comparison yields a p-value of approximately zero, which means that we can reject the null hypothesis that changing from and 85 calorie diet to a 50 calorie diet has no impact on lifespan.

```
RankSumTest[lifespans[[3]], lifespans[[4]],
 TwoSided → True,
 ExactDistribution → True]
```

Rank Sum Statistic:	TwoSided P Value:	Sampling Distribution:
4512.5	0.879811	Exact

Question 2 corresponds to testing for group differences between N/R50 group and the R/R50 group because both groups had 50 calorie limits, but one started before weaning and one started after weaning. As we can see from the rank sum test, the comparison yields a p-value of approximately 0.879, so we can not reject the null hypothesis that starting a 50 calorie diet after weaning has a different impact on lifespan than starting the same diet before weaning.

```
RankSumTest[lifespans[[4]], lifespans[[5]],
 TwoSided → True,
 ExactDistribution → True]
```

Rank Sum Statistic:	TwoSided P Value:	Sampling Distribution:
3613.	0.00865283	Exact

Question 3 corresponding to testing for group differences between the R/R50 group and the R/R50L group because both groups had 50 calorie diets before and after weaning, but one also had a protein reduction in their diets after weaning. As we can see from the rank sum test above, the comparison yielded a p-value of approximately 0.0086, so we can reject the null hypothesis that a reduction in protein had no impact on lifespans.

```
RankSumTest[lifespans[[3]], lifespans[[6]],
 TwoSided → True,
 ExactDistribution → True]
```

Rank Sum Statistic:	TwoSided P Value:	Sampling Distribution:
4238.5	0.0385078	Exact

Question 4 corresponds to testing for group differences between the N/R50 group and N/R40 group because both groups had no restrictions before weaning, and had 50 and 40 calorie restrictions after weaning, respectively (this test could also be performed between the rats restricted to 40 and the rats restricted to 85, but from the summary statistics it appears that calorie restrictions led to longer lifespans, and we already observed a difference between a restriction from 85 calories to 50 calories, so it is safe to assume that the 40-85 comparison would yield a similarly robust result). As we can see from the rank sum test above, the comparison yields a p-value of 0.038, so we can reject the null hypothesis that changing a rat from a 50 calorie diet to a 40 calorie diet would have no impact.

Final Problem 6: Lifespan Study

From the box plots in part (a), it appears the restricting rats to fewer calories generally increased their lifespans. Although the variation and amount of outliers did not look so sever in the data, a normal probability plot made it seem like the data was not normally distributed, so I opted to use non-parametric methods for my statistical analysis. Using the Kruskal-Wallis test, we could safely conclude that there were differences in lifespans between the separate groups. Then, by using rank sum tests, we could observe that reducing rats from 85 calorie to 50, reductions in protein, and reducing rats from a 50 calorie per week to a 40 calorie per week diet all had significant impacts on their lifespans. From the summary statistics, it appears that both caloric reductions yielded an increase in lifespans, and the reduction in protein yielded a decrease in lifespans. Only pre-weaning dietary restrictions did not have any impact.

Inclusion of the NP group would be nice to have if our direct comparison of groups had no initial effects. For instance, if it turned out that if restricting rats from 85 calories to 50 calories per week, we could also test whether calorie restrictions had any impact by comparing the group with 85 calories and the group with no restrictions. Furthermore, each group can be compared to the rats with no restrictions to test whether the restrictions make any differences compared to unrestricted rats. This could help to increase the robustness of the tests.