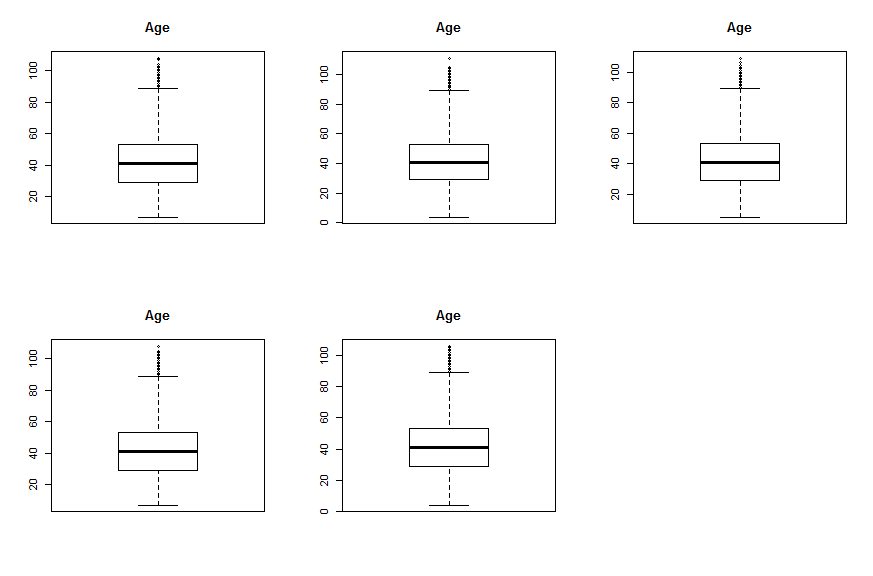
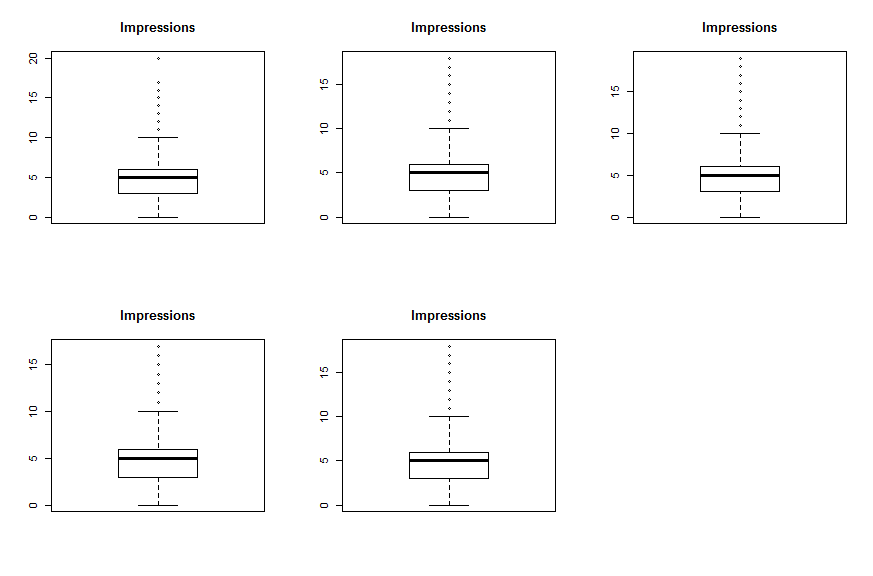
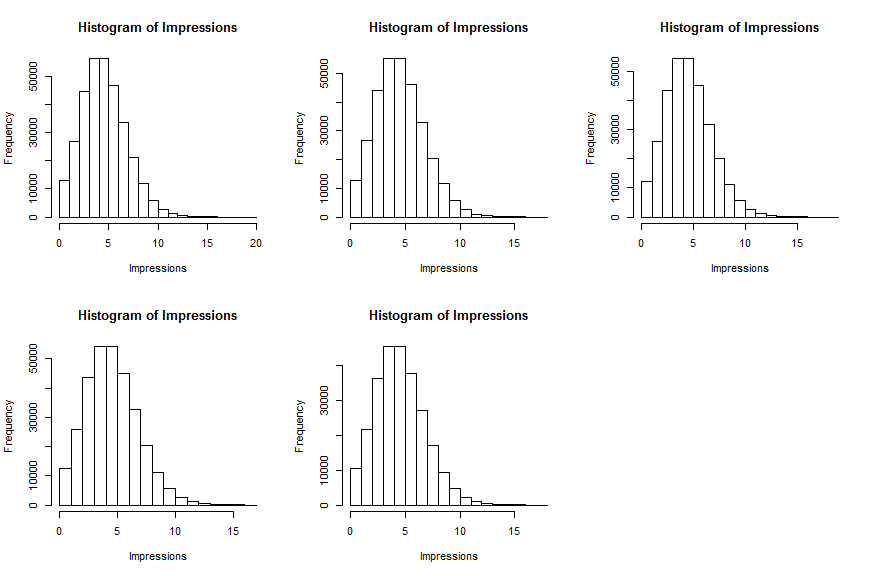
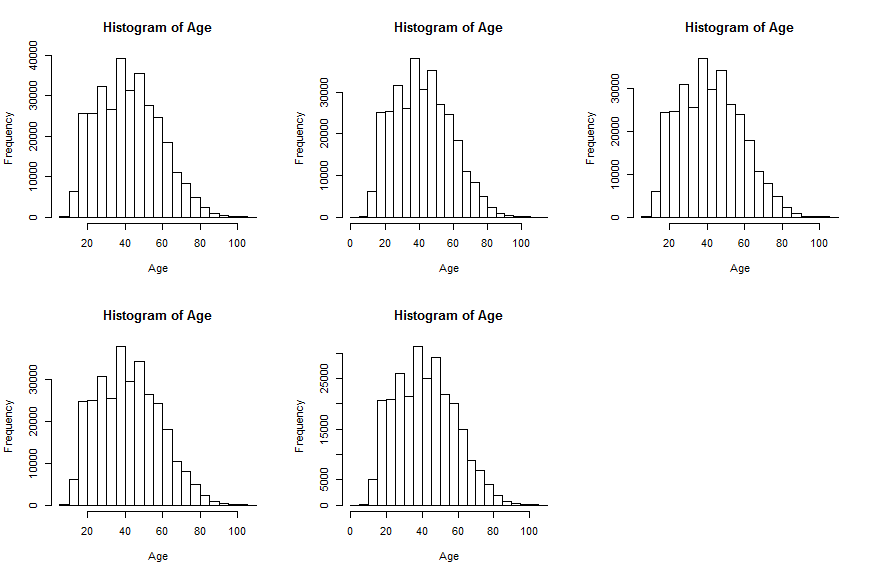
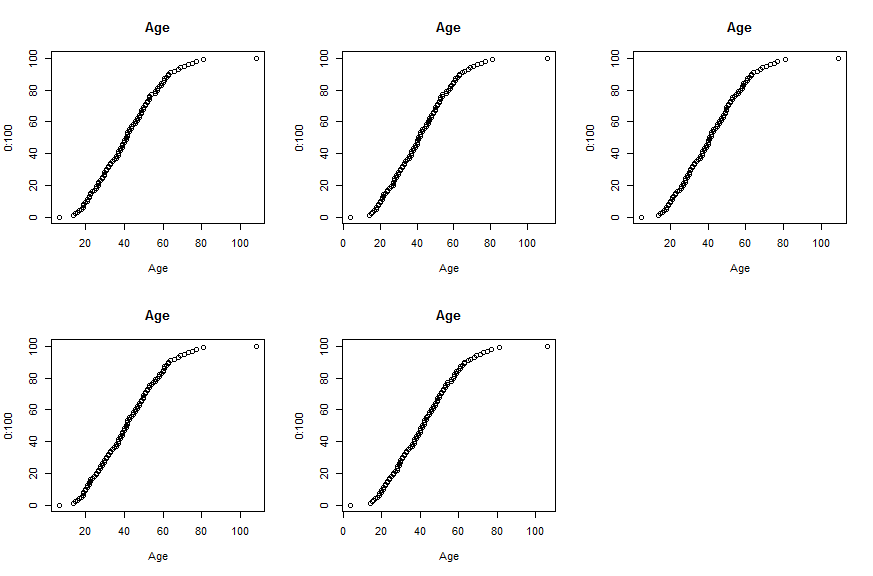
1. A. The distributions of Age and Impression appear stationary across datasets. The mean age is approximately 40 and the median impressions is 5. There are fairly large inter-quintile ranges for all distributions.

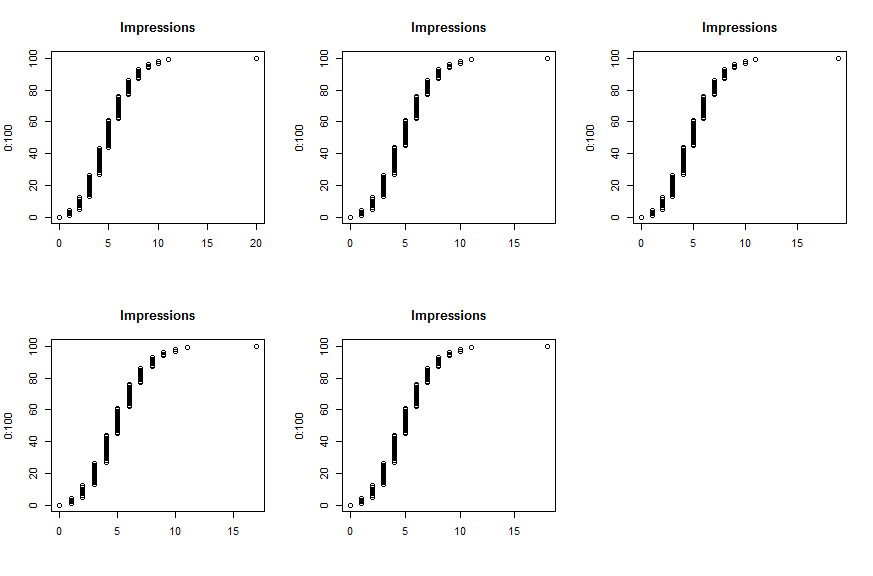


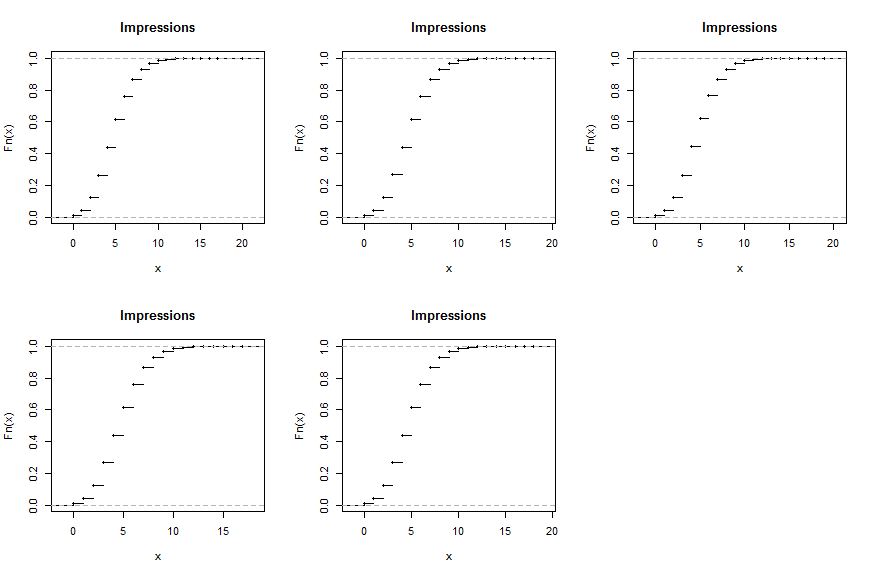
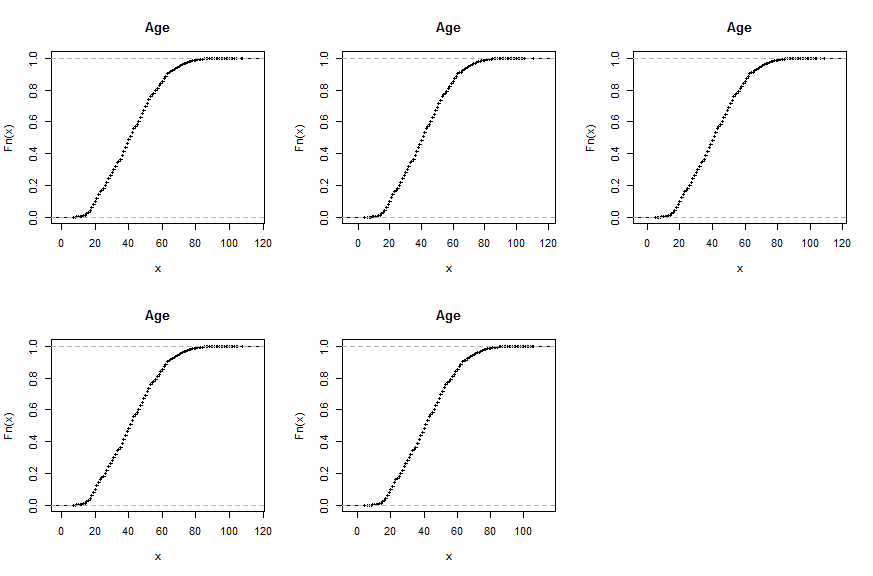
B. These look almost normal, but lean to the left. The distribution of Age looks almost like two distributions. The distribution of Impressions appears constant at this number of bins across time. 



C. Age and impressions look like light-tailed distributions. There is enough data for Age to appear smooth, though impressions are over a small enough range that it looks discrete. There is enough data for both to stabilize.







D. The null hypothesis of the Shapiro-Wilk test is that the population is normally distributed. The tests are all significant at greater than 95% significance. We can reject the null hypothesis that the populations are normally distributed. This means we can infer that the distributions are not normally distributed.

> exb(dss, stest, "Impressions")

[[1]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9723, p-value < 2.2e-16

[[2]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9689, p-value < 2.2e-16

[[3]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9718, p-value < 2.2e-16

[[4]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9718, p-value < 2.2e-16

[[5]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9696, p-value < 2.2e-16

> ex(dss, stest, "Age")

[[1]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9805, p-value < 2.2e-16

[[2]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9789, p-value < 2.2e-16

[[3]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9792, p-value < 2.2e-16

[[4]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9779, p-value < 2.2e-16

[[5]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

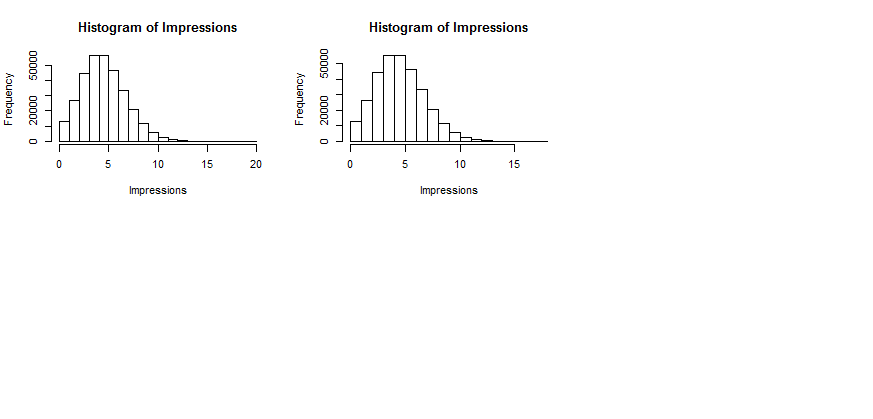
W = 0.981, p-value < 2.2e-16

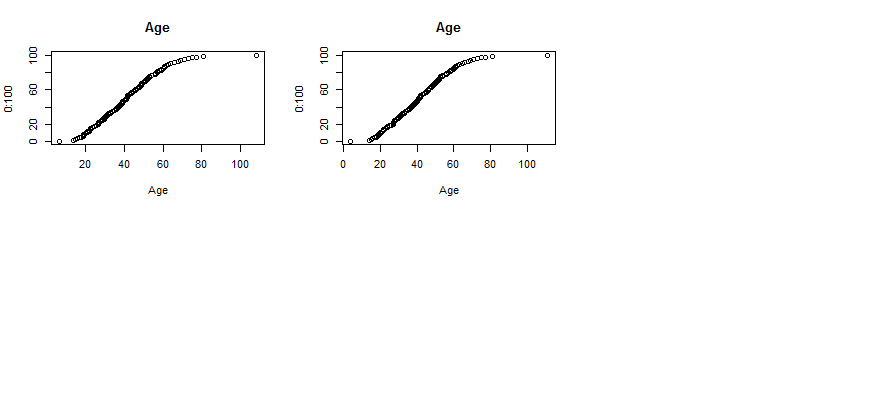
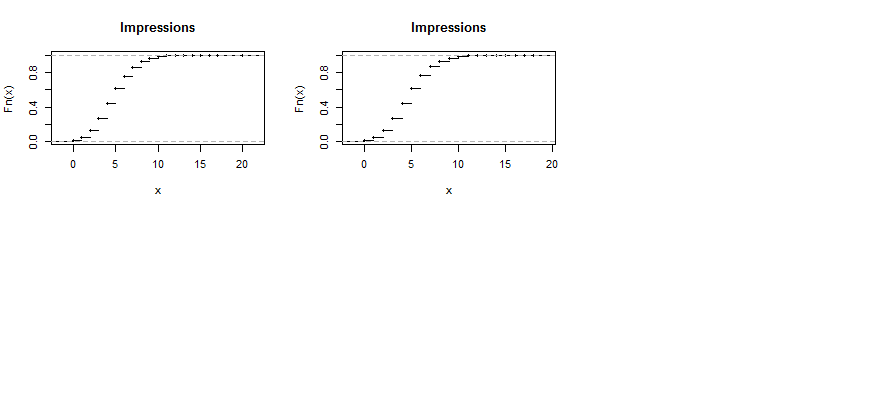
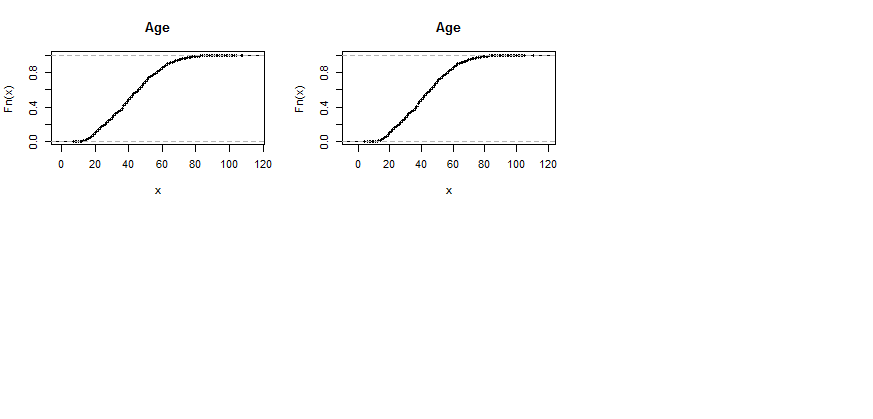
E. From age, we can see that the New York Times has a very diverse audience and has a reach across age groups. There are certainly problems with it though, because age should not be 0.

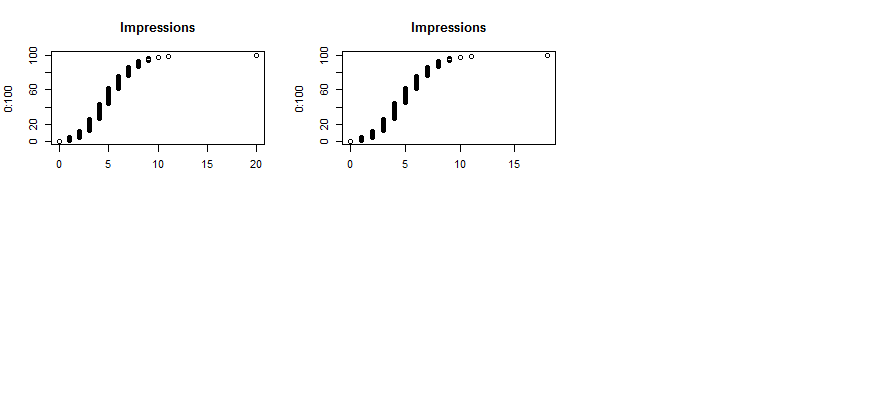
1. Extra Credit

Filtering on Signed\_In so only registered users are counted.









[[1]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9804, p-value < 2.2e-16

[[2]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9798, p-value < 2.2e-16

> exb(filtered, stest, "Impressions")

[[1]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9697, p-value < 2.2e-16

[[2]]

Shapiro-Wilk normality test

data: sample(e, size = 5000, replace = TRUE)

W = 0.9692, p-value < 2.2e-16

Filtering by signed in users yields similar results. This would suggest that impressions are independent of whether or not a reader is signed in. The same goes for age, suggesting that age is independent of registering as well.