Inference for Discrete Time Markov Chains

[Avery and Henderson (1999)](#) discusses the use of Markov chains in modeling DNA sequences (in the preproglucacon protein). The data set `preproglucacon` in the `markovchain` package contains related data.

```
library(markovchain)

data("preproglucacon", package = "markovchain")

x = preproglucacon$preproglucacon

head(x)
```

```
[1] "G" "T" "A" "T" "T" "A"
```

1. Use the data to estimate the transition matrix.
2. Find and interpret a 95% confidence interval for $p(A, C)$.
3. Find the stationary distribution that corresponds to the estimated transition matrix.
4. Find the observed relative frequency of each state in the data. How do the observed relative frequencies compare to the stationary distribution?
5. Create an appropriate plot to assess if the observed sequence can be reasonably considered an observation from a (first order) Markov chain.
6. (Optional.) Use the `markovchain` package to test if the observed sequence can be reasonably considered an observation from a (first order) Markov chain