36-201: Statistical Reasoning

Computer Lab 14: Time Series Data and Transformations
Solutions

Example I

Question #1: In March 1986 there were 97,987,000 employed workers. In March 1987 there were 100,202,000.

Question #2: Figure 1 shows the time series. The cycle is about a year long.

♣ **Question #3: There is a two-hump seasonal variation each year, with the valley between the humps centered at about June or July of each year. The rise from January to June could be due to seasonal construction work being available; the rise at the end of the year could be due to seasonal sales jobs. (Other plausible explanations also accepted).**

Question #4: The trend is increasing. Overall more people are employed each year.

Question #5: Figure 2 shows the boxplots. A smooth line through the medians of the boxplots would be roughly S-shaped, increasing steadily from 1986 to 1990 and from 1992 to 1996, but with some flatness or even decrease from 1990 to 1992.

Question #6: The 10th boxplot can be ignored since it is based on only one data value (January 1996). Of the other nine, most are fairly symmetric or slightly skewed left (longer left tail than right tail), and none have outliers. The variability is greatest where the employment is steadily increasing, and it gets less in the years 1990 and 1991 where the number of employed workers is actually decreasing.

Question #7: As observed in previous questions, there is a decrease in the number of employed workers from 1990 to 1991. Employment was fairly flat from 1991 to 1992, with the main difference between the two years being the amount of variability in employment throughout the year.

♣ **Question #8: Clinton was running against George Bush, who was President from 1988 to 1992. Clinton, with Carville's advice, was trying to hang a poor economy on Bush; the weakness of the economy is indicated in the time series plot by the drop in employment from 1990 to 1991 and 1992.**

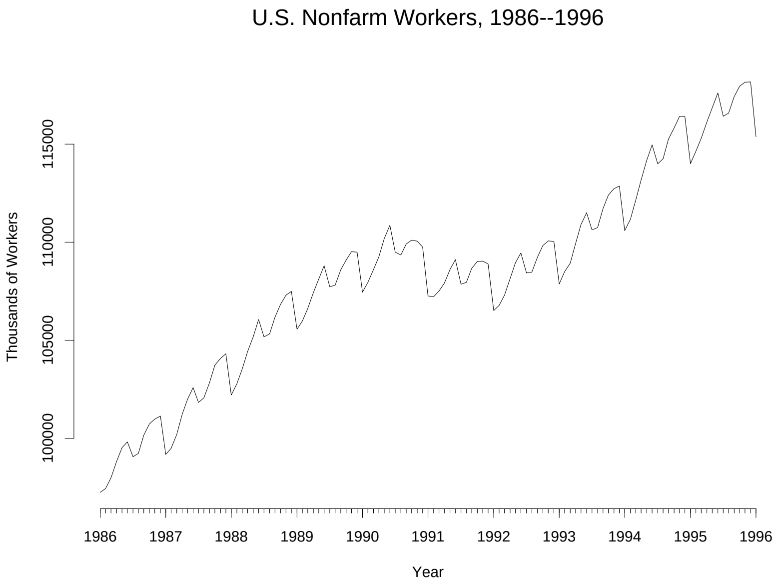


Figure 1: Monthly Employment Totals for a Decade

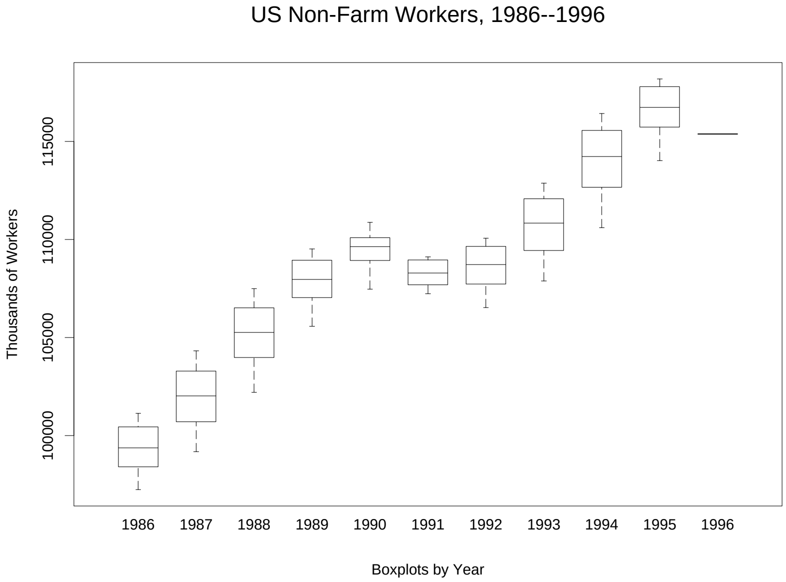
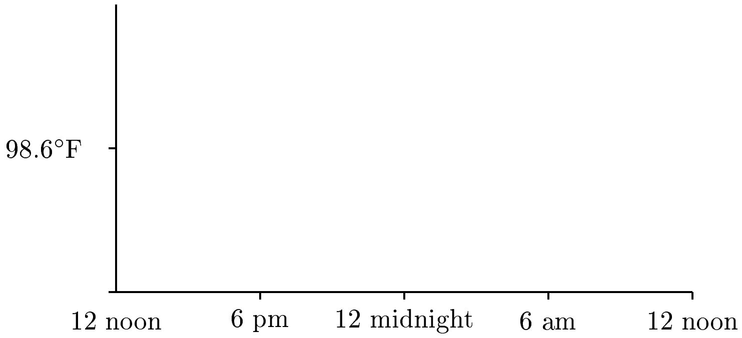


Figure 2: Monthly Employment Totals for a Decade

Example II

Question #9: Here is my guess.



Question #10: Figure 3 on the next page shows the time series.

♣ **Question #11: There does not appear to be a trend in body temperature. However there is a very clear cycle in body temperature. The minimum temperature occurs around 3 am and the maximum occurs around 2 pm. The temperature starts to rise around the time you start waking up in the morning. This pattern is repeated every 24 hours and can be seen clearly in Figure 3.**

Question #12: The MIN is about 36.75°C and the MAX is about 38.5°C. The range is about 38.5–36.75=1.75°C.

Question #13: Using **Descriptive Statistics** in MINITAB we get that the mean is 37.53; the standard deviation is 0.3713.

Question #14: The boxplot is fairly symmetric, with no outliers. You can't see any of the time-dependent features with a single boxplot, since time is not even included in the plot.

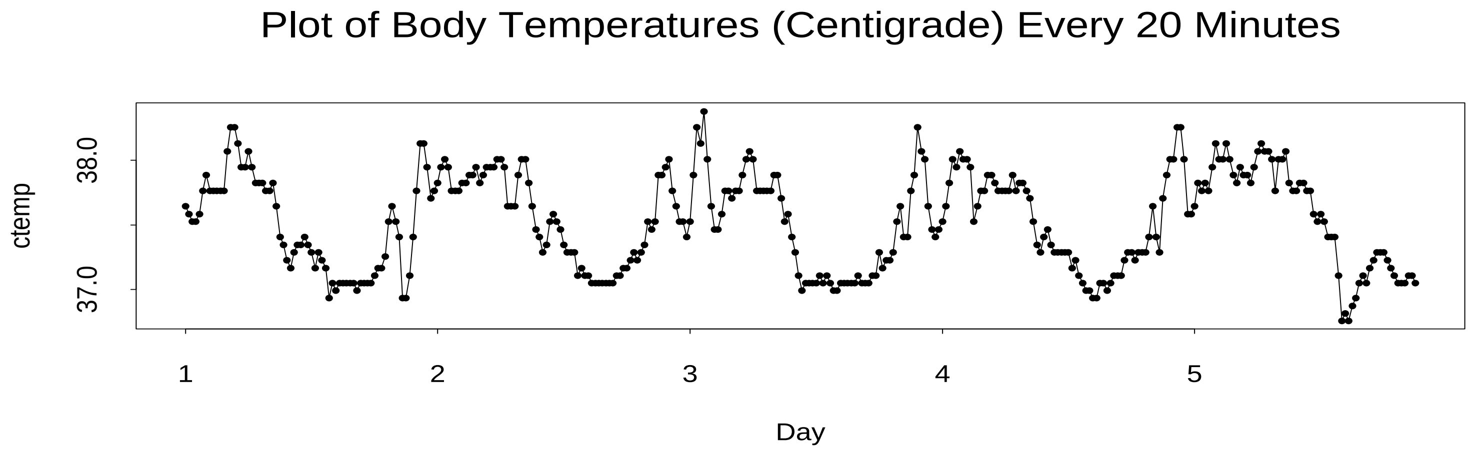
♣ **Question #15: The time series plot of the temperature data in °F is also shown in Figure 3. Its features are identical to the time series using °C.** *Thus the shapes of graphical plots are unaffected by changes in location and scale of the data.*

Question #16: The new mean is 99.55 and the new SD is 0.6684. Note that
(new mean) = 1.8*(old mean) + 32; and (new SD) = 1.8*(old SD)
This is an example of the following facts, which are *always* true:

If you add a constant to the data, you also add it to $\overline{x}$, but not to SD_x: If you add 3 to every observation, the new mean becomes $\overline{x} + 3$ but the new SD is exactly the same as the old SD_x.

If you multiply the data by a constant, you multiply both $\overline{x}$ and SD_x: If you multiply every observation by 5, the new mean becomes $5 \cdot \overline{x}$ and the new standard deviation becomes $5 \cdot SD_x$.

Question #17: This is a freebie.



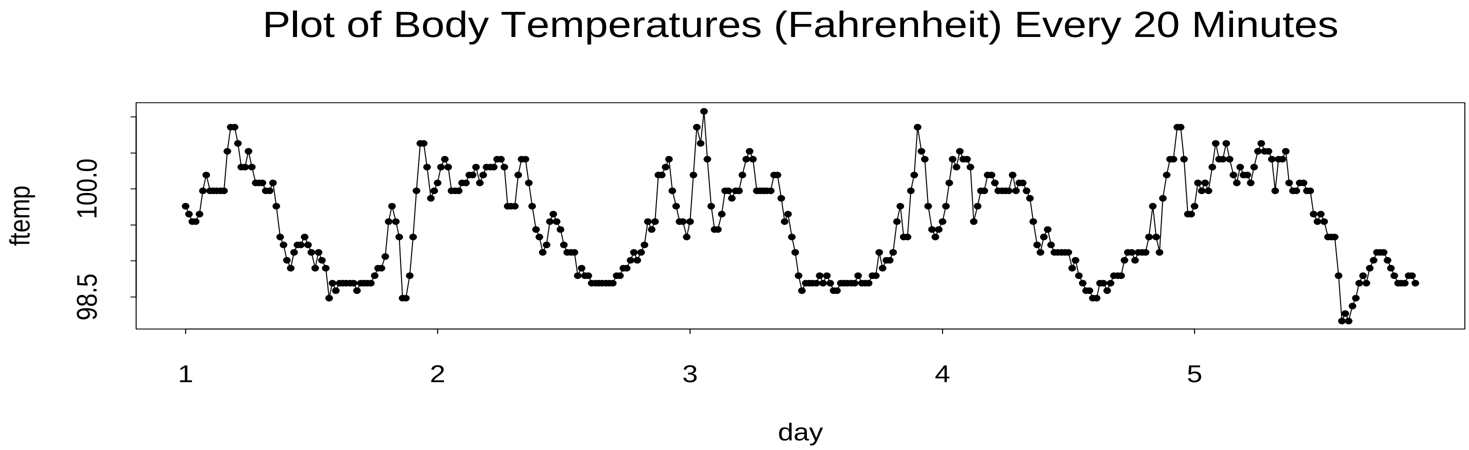


Figure 3: Body Temperature Taken Every 20 Minutes