1. Please see the code for details. This is a one-way ANOVA situation. We do the ANOVA and then study the diagnostic plots. The variance is not constant, and the residuals are not normally distributed. After a logarithmic data transformation the results look a lot better! The ANOVA table is:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>4</td>
<td>4.62</td>
<td>1.15</td>
<td>4.29</td>
<td>0.0041</td>
</tr>
<tr>
<td>Residuals</td>
<td>59</td>
<td>15.89</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We get a P-value of 0.004, indicating strong evidence for differences among the cancer types.

[ 4 points ]

2. Please see the code for more details.

(a) A stripchart for example is a good way of showing the data.

(b) The null hypothesis is that all group means are the same, $H_0: \mu_1 = \cdots = \mu_5$, versus the alternative that they are not all the same. A one-way ANOVA yields a F-statistic of 3.76, and a p-value of 0.009 (4 and 60 degrees of freedom), so we conclude that indeed there are differences in the stem lengths of daffodils from the different sites.

(c) Both approaches (assuring a 5% family-wise error rate) detect significant differences between the open area and the areas north and south of the building, but otherwise, do not indicate differences at this significance level.

(d) The Kruskal-Wallis test yields a p-value of 0.005, and thus, the differences between sites are more significant than those from the parametric one-way ANOVA.

(e) Both tests are still significant, with the one-way ANOVA becoming way more significant (0.0009 versus 0.009 before removing the smallest observation). This appears to be largely due to the within-group variance being much smaller after removing the smallest observation (2192.4 versus 3034.3). The Kruskal-Wallace test yields about the same significance (0.003 versus 0.005 before).

(f) We do not detect any significant differences between any of the four sites around the building (assuring a 5% family-wise error rate via Bonferroni), but do now detect significant differences between all of the sites around the building when comparing the stem lengths to those in the open area nearby (Tukey’s HSD yield the same info, btw).

(g) Overall, it appears that the stem lengths do not differ much when comparing the daffodils from the four sides of the building, but the stem lengths appear to be longer compared to the daffodils from the open area. Maybe they daffodils have to “stretch” some more around the building to catch the sunlight? There was one very short daffodil collected on the west side of the building that had a fair amount of influence on the statistical inference: the differences between the stem lengths from the open area and those from the west and east side of the building were not statistically significant (though close!) after multiple comparisons correction when including that flower.

[ 6 points ]
3. Since we are interested in this particular treatment, group is a fixed effect. The mice are nested in the groups, and are considered a random effect, since we are not interested in these particular mice, but have to account for the possible biological variation. The factor group has two levels (treatment/control) and therefore one degree of freedom. The factor mouse is nested within group, and has $2 \times (5-1) = 8$ degrees of freedom. For the error we have $2 \times 5 \times (2-1) = 10$ degrees of freedom. The respective mean squares are therefore 4.37, 5.43, and 0.79. The F statistics are $4.37/5.43 = 0.80$ for the factor groups, and $5.43/0.79 = 6.87$ for the factor mouse. The p-values are 0.396 and 0.003, respectively. There is no evidence for a treatment effect, 0.80 is not an extreme draw from a F distribution with 1 and 8 degrees of freedom. However, there seems to be biological variability between mice, 6.87 is a somewhat extreme draw from a F distribution with 8 and 10 degrees of freedom.

[ 3 points ]

4. Please see the code for the details of the simulation. When the means are 3, 4, 5 and 6 respectively, the estimate I get is 81.4% power. Since 8,140 out of the 10,000 iterations yielded a p-value less than 0.05, the 95% confidence interval for the power can be calculated with `binom.test()`, which gives (0.806;0.821). Note that a rough 95% confidence interval for the power can be calculated as $81.4\% \pm 1\%$ (i.e., $\hat{p} \pm 1/\sqrt{n}$). When the means are 3, 3, 6 and 6 respectively, we expect more power because the squared treatment effects ($\alpha_i$) are larger, and therefore the expected means squares for the treatment effects are larger. The simulation shows 98% power.

[ 6 points ]