Healthy Options 2015 Report
Math 107-A/B: Applied Statistics, Gettysburg College
Professor: Kim Spayd

In the Spring 2016 semester, students in two sections of Applied Statistics analyzed provided data from the 2015 Health Options program. This report addresses seven questions posed to the students.

1. **What types of vendors do Healthy Options, EBT and Debit customers shop at?**

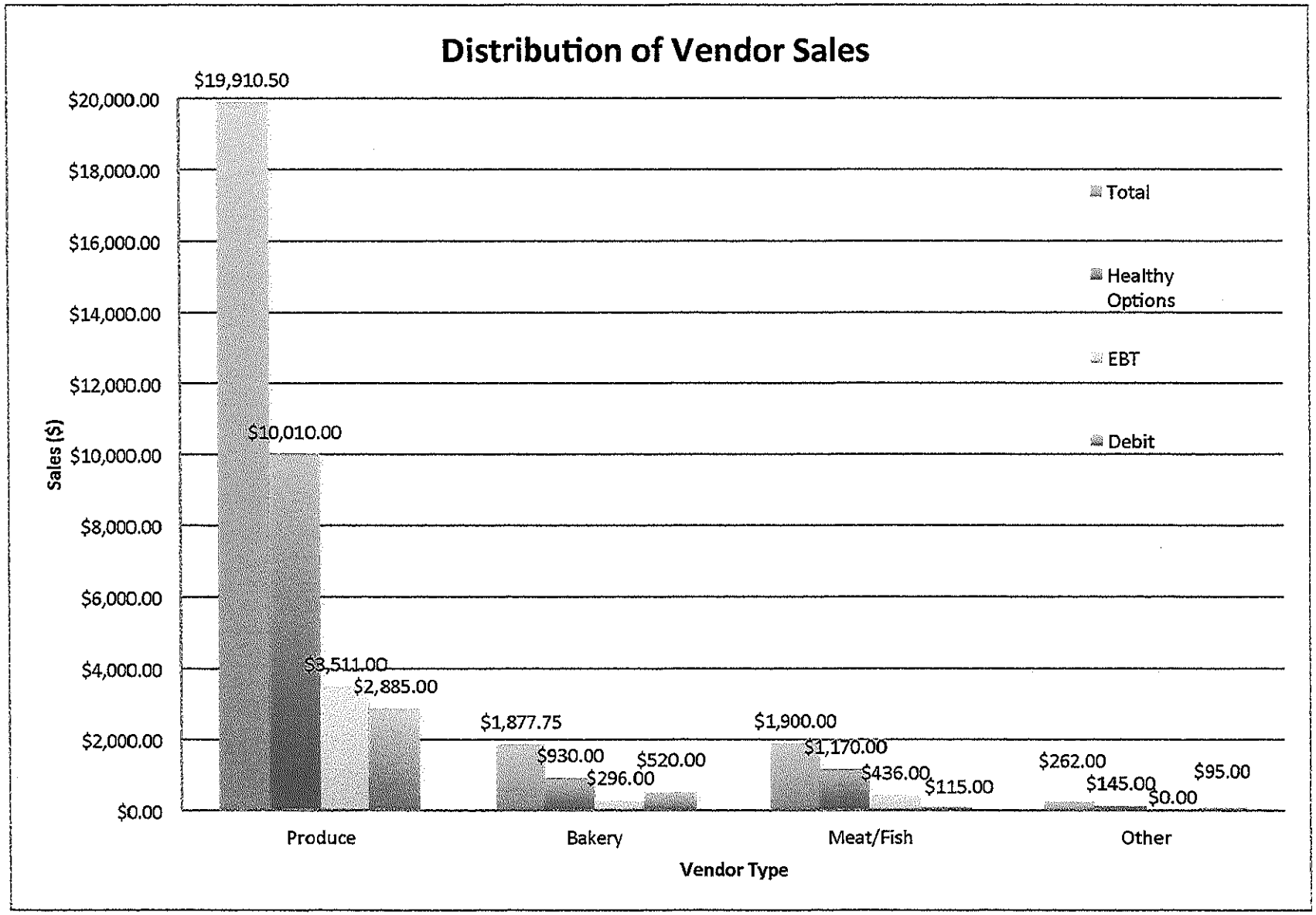
The attached bar graph shows that the vast majority of sales occur at produce vendors. It is noteworthy that the 6 produce vendors (approximately 43% of vendors) account for approximately 83% of all Healthy Options, EBT, Debit, FMNP tokens, Market Bucks and Kids Bucks receipts. Among Healthy Options sales, produce receipts greatly outnumber non-produce purchases. The produce sales from Healthy Options participants are nearly triple EBT sales and Debit sales, separately.

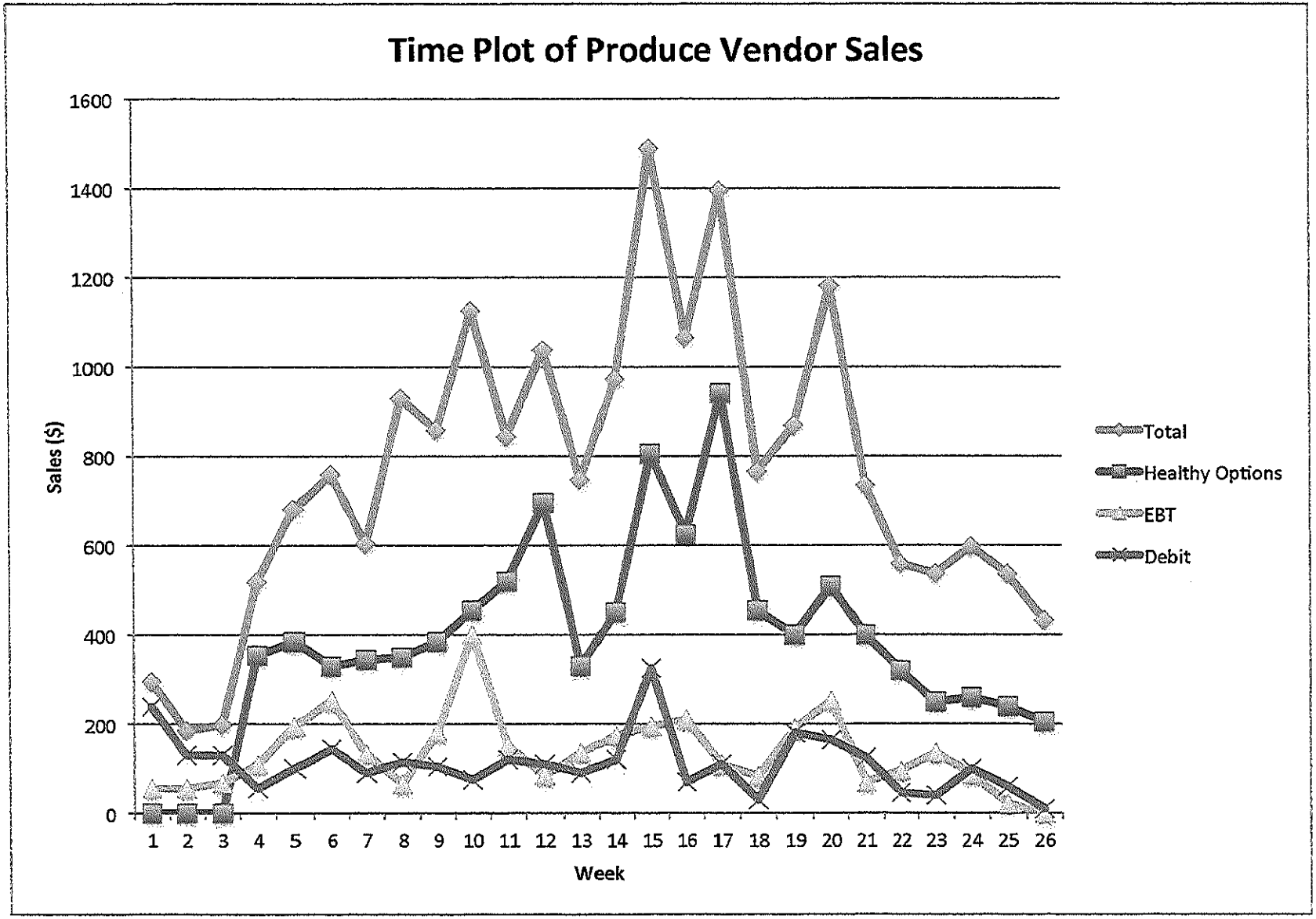
2. **Does spending on various items change at different times of the season? Does that vary for shoppers using the different outreach programs?**

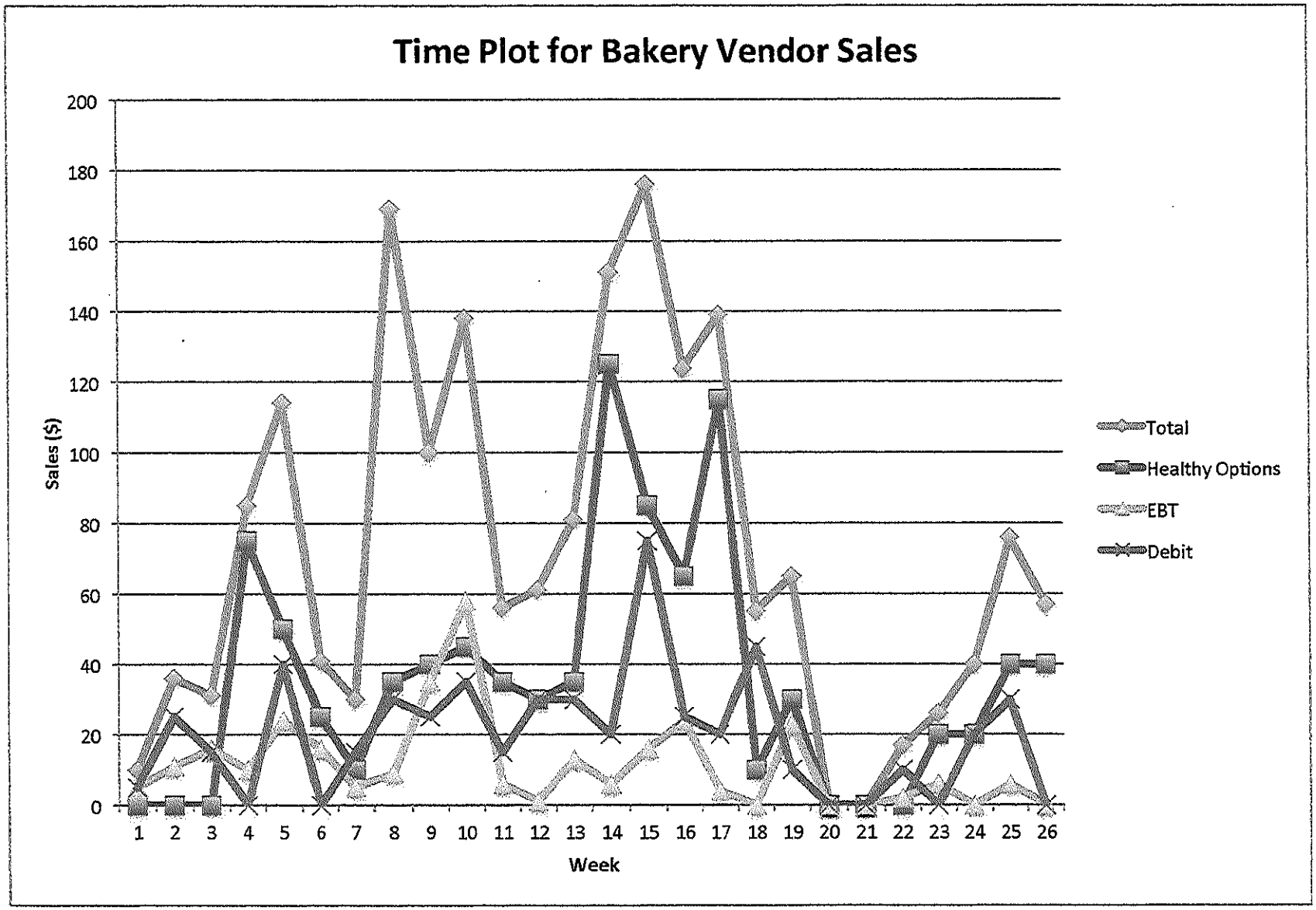
From the four attached time plots, there is some evidence to support this claim in produce and bakery sales from Healthy Options sales but not particularly for meat/fish and other (soap and jewelry) vendors. It appears that there is a gradual rise through the first two months into a general peak in the middle of July and through August, then a gradual decline through the last two months of the market. The trend can be seen also with the average sales per month in each program. See the table below.

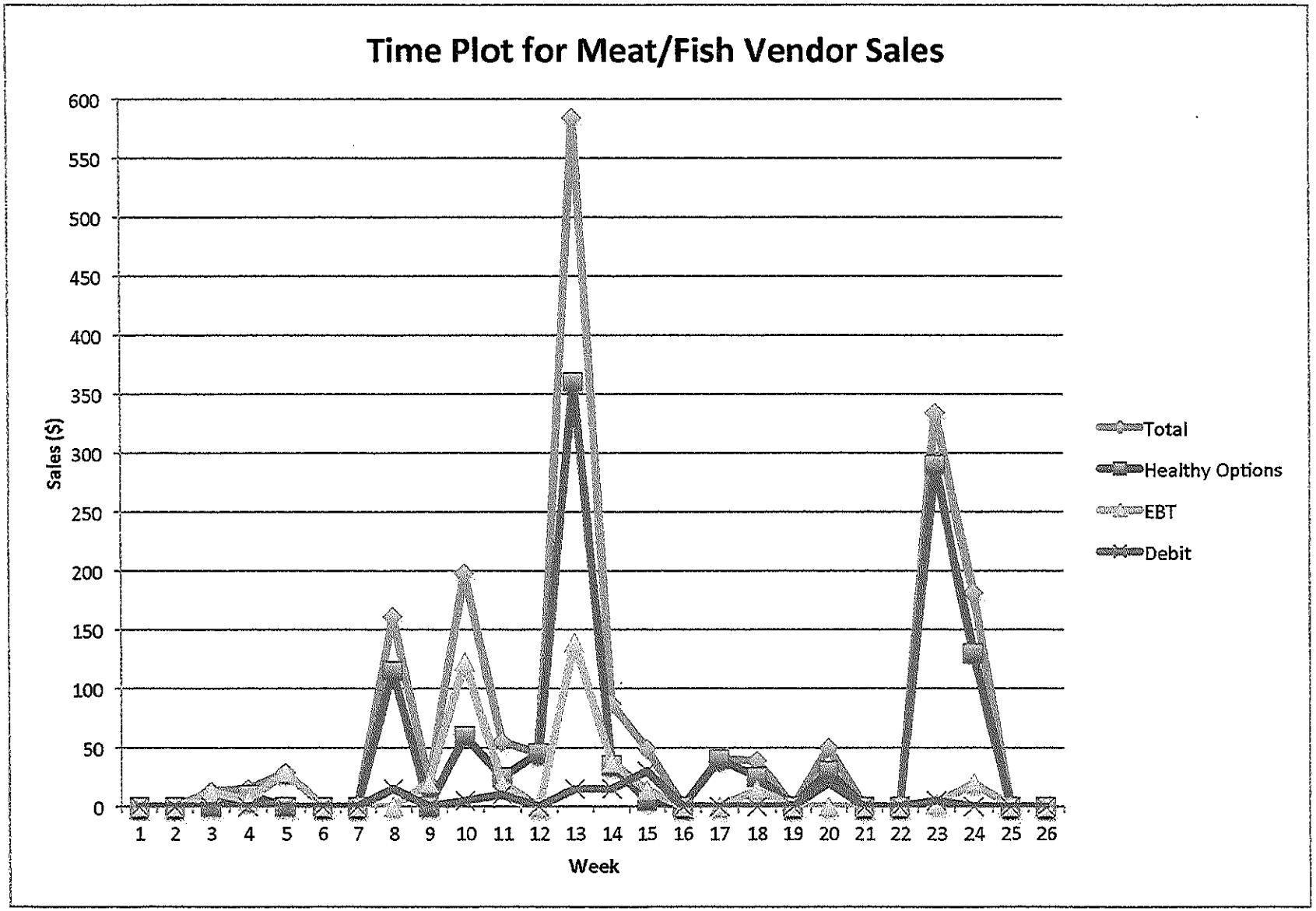
	May	June	July	August	September	October
Healthy Options	$355.00	$352.50	$477.00	$705.00	$441.25	$255.00
EBT	$71.75	$161.50	$190.60	$172.25	$149.00	$85.00
Debit	$138.75	$112.50	$100.00	$156.25	$125.00	$51.00

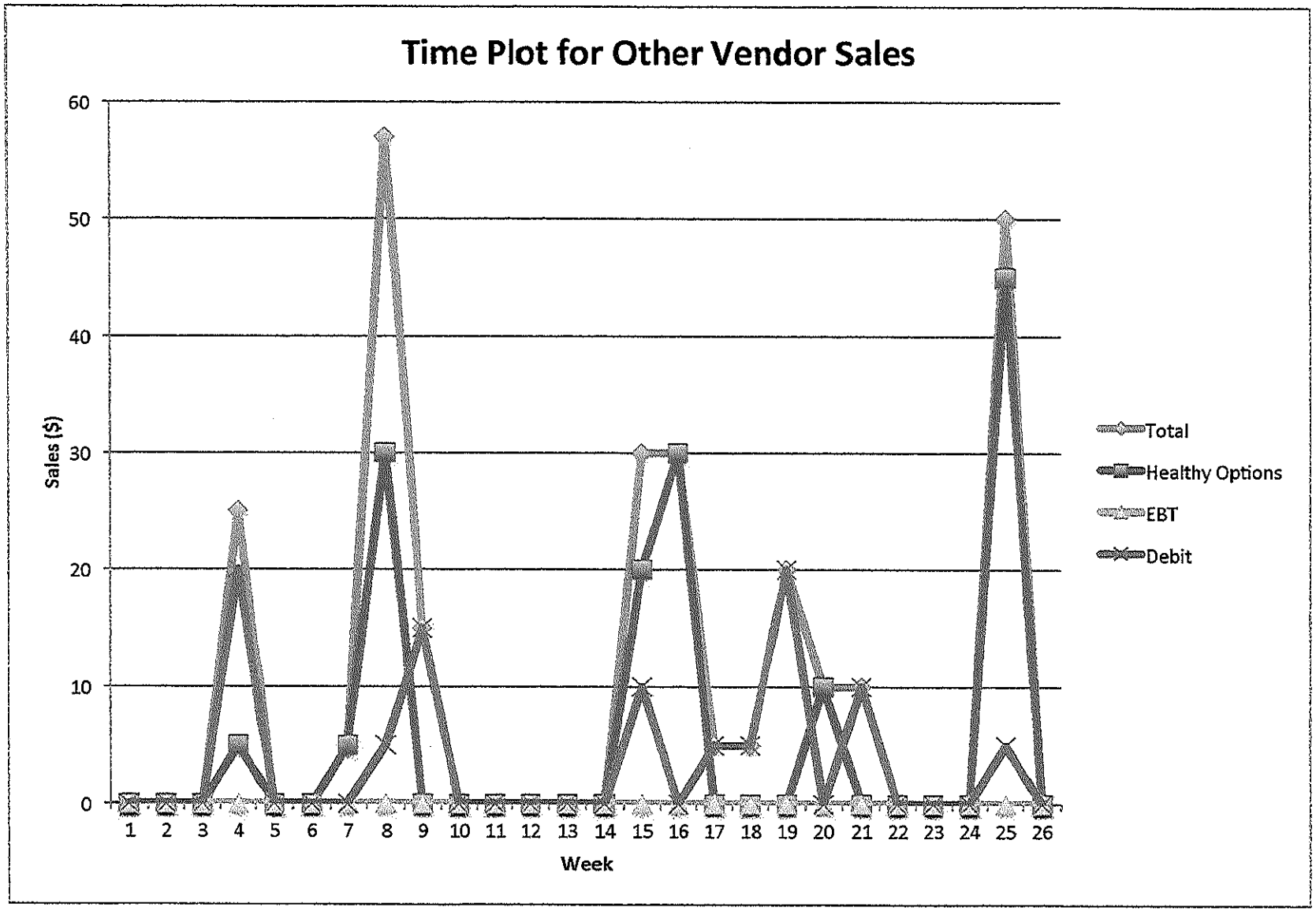
To further investigate the seasonal impact on produce receipts, we would have liked to perform separate analyses of variance (ANOVA) for monthly average Healthy Options, EBT and Debit sales. However, the standard deviations of each month's receipts do not allow for this. The rule of thumb typically used, to ensure that the ANOVA test's results are approximately accurate, is that the largest standard deviation should be no larger than twice the smallest standard. Even when the data are reorganized into three groups, each one covering two-month periods, the standard deviations do not follow the rule of thumb. Thus we regretfully cannot perform any statistical inference to rigorously test whether there is a significant impact on spending by month or season.


Distribution of Vendor Sales
$20,000.00
$18,000.00
$16,000.00
$14,000.00
$12,000.00
$10,000.00
$8,000.00
$6,000.00
$4,000.00
$2,000.00
$0.00
Sales ($)
$19,910.50
$10,010.00
$3,511.00
$2,885.00
$1,877.75
$930.00
$296.00
$520.00
$1,900.00
$1,170.00
$436.00
$115.00
$262.00
$145.00
$0.00
$95.00
Total
Healthy Options
EBT
Debit
Produce
Bakery
Meat/Fish
Other
Vendor Type


Time Plot of Produce Vendor Sales
1600
1400
1200
1000
800
600
400
200
0
Sales ($)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Week
Total
Healthy Options
EBT
Debit

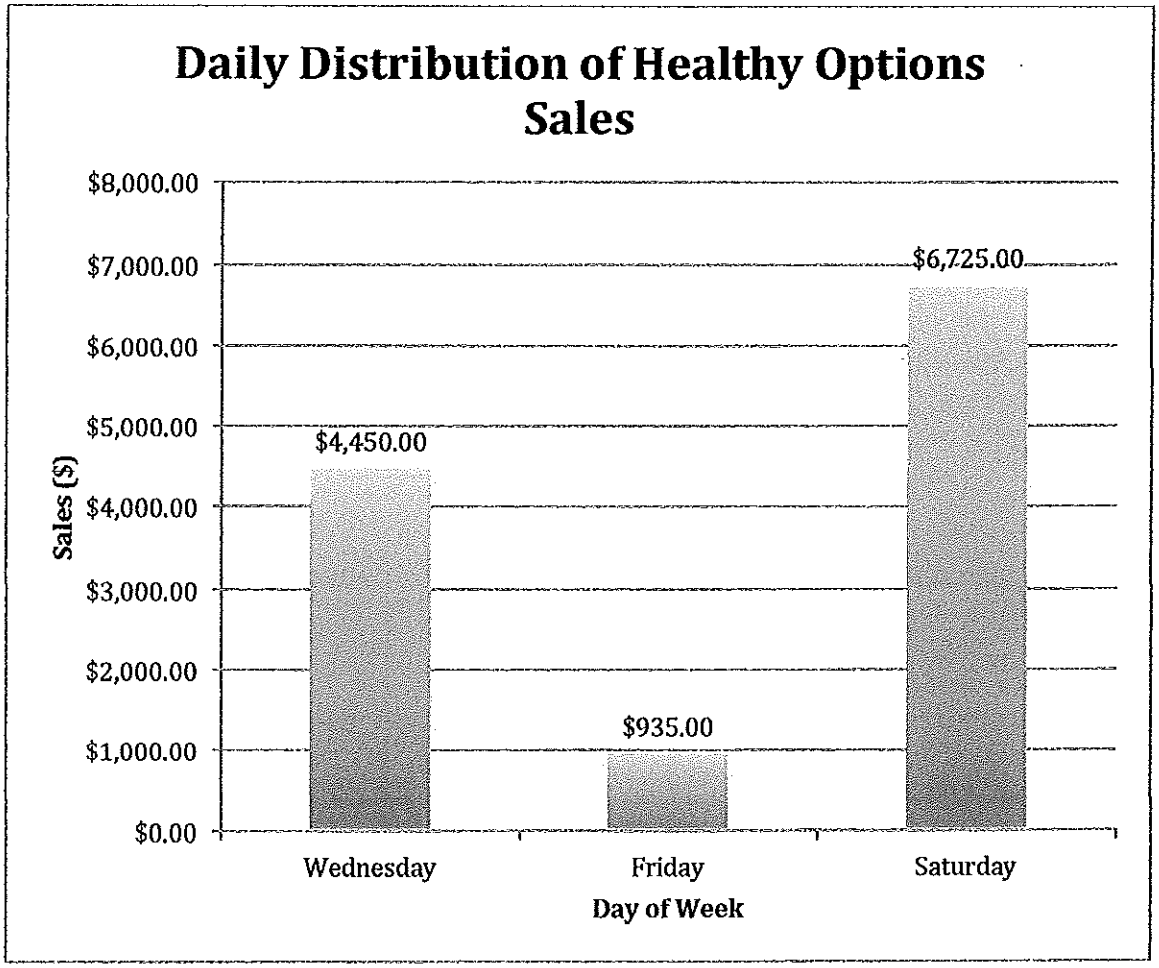

Time Plot for Bakery Vendor Sales
Sales ($)
200
180
160
140
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Week
Total
Healthy Options
EBT
Debit


Time Plot for Meat/Fish Vendor Sales
600
550
500
450
400
350
300
250
200
150
100
50
0
Sales ($)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Week
Total
Healthy Options
EBT
Debit


Time Plot for Other Vendor Sales
60
50
40
30
20
10
0
Sales ($)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Week
Total
Healthy Options
EBT
Debit

3. **Does the day of the market impact spending patterns of Healthy Options participants?**

Here, our suspicion at first is that total Saturday receipts would be significantly larger than those on Wednesday and Friday since work demands might prohibit people from visiting the market on weekdays. The bar graph below lends support to this hypothesis; however it was surprising to see the large difference between Wednesday and Friday receipts.



To be rigorous in discerning a statistically significant difference between Wednesday and Saturday sales, we calculate both a 90% confidence interval and a 95% confidence interval to estimate the difference between the average amount spent on Saturdays and Wednesdays:

- 90% confidence interval: ($11.42, $181.92)
- 95% confidence interval: (-$6.44, $199.77)

In the first case, the 90% confidence interval tells us that we can expect, on average, Saturday receipts to be between $11.42 and $181.92 more than Wednesday receipts. *Thus, with 90% accuracy, we can say that the day of the week does have a statistically significance impact on spending.* However, if we increase accuracy of the method to 95%, we lose statistical significance. This comes from the -$6.44 at the

lower end of the 95% confidence interval, which is interpreted as a possible average of $6.44 more on Wednesdays that Saturdays. With $0 included in the 95% confidence interval, we see that the method allows for average sales to be the same on the two days in question. *In this case, the day of the week does not have a statistically significant impact on spending.*

For both intervals, we use the Student t distribution with 20 degrees of freedom (based on the 21 Wednesday observations). While the data is not from an ideally random sample of all Adams County residents who fall in the food gap, we felt that the Healthy Options participants would resemble the larger population of interest well enough for the inference methods to apply reasonably well. Additionally, since the combined number of observations from Wednesdays and Saturdays is relatively large (43), the confidence interval estimates are quite robust.

4. **Does a participant's town of residence impact Healthy Options voucher usage?**

After some discussion, the students hypothesized that Gettysburg residents would be more likely to visit the market because of the proximity. To test this, many students chose to compare the percentages of vouchers used between Gettysburg residents and non-Gettysburg residents. There are six Healthy Options participants who are left out of this analysis because no residence information was available. Of the remaining 68 participants, 29 are Gettysburg residents and 39 are not. The 29 Gettysburg residents used 92.72% of available vouchers (968 of 1044). The 39 non-Gettysburg residents used 93.02% of available vouchers (1306 of 1404). Based on these values, the test statistic is calculated to be z=-0.047557, using the Normal distribution. The associated P-value is 0.518965, too large to give statistical significance to town of residence. *In conclusion, then, there is no evidence from this data that the town of residence has an impact on voucher usage.*

We are assuming that the participants in the two groups are independent; that is, we are assuming that spending in one group will have no effect on spending in the other group. Given the social component of the Healthy Options program, this may not be a valid assumption. For example, it is possible that participants from different towns would plan to meet at the market, thereby creating dependence of voucher usage between the two groups. If we disregard this issue, then the demographics of the participants and sample sizes allow for high accuracy of the results.

5. **Does weather impact Healthy Options voucher usage?**

Based on the provided weather notes, the students classified the weather on each day the market was open as either "nice" or "not nice." A particular day is classified as "not nice" if any key words in the weather notes give negative connotations for market attendance, such as "rain," "humid," and "cold." Otherwise, a particular day was classified as "nice" if key words such as "perfect," "sunny," or "beautiful" appear in the weather notes. Of the 53 days with weather qualifiers, 32 are deemed "nice",

with an average of $220.16 and a standard deviation of $208.10 being spent at the market. Of the 23 days that are deemed "not nice," the average amount spent is $154.35 with a standard deviation of $124.75. While the averages are markedly different, the standard deviations are large enough to leave the statistical significance of weather in question. To address it, we construct a 90% confidence interval for the difference in the average amount spent on "nice" and "not nice" days: ($-11.56, $143.17). As with the third question regarding day of the week, this interval includes $0, allowing for the average amounts spent in both types of weather to be the same. *Thus we cannot say that weather plays a statistically significant role in voucher usage.* There is no need to construct higher-level intervals since they will only increase in width and will all include $0.

The possible dependence between the two types of weather is an issue again, as in the question regarding town of residence. Disregarding it, all other assumptions for robust inference hold.

6. Does a Healthy Options participant's primary language affect spending?

To answer this question, we take the same approach as with town residence. The participants are grouped into "English-speaking" and "Non-English-speaking" categories. The second group includes all Spanish-speaking participants as well as the single Haitian Creole-speaking participant. If the language preference includes both English and Spanish for any participant, that person is classified as English-speaking. The students thought this would be a natural designation given their suspicion that a language barrier would discourage participants from visiting the market. Thus the operating hypothesis is that there is no difference in the proportion of vouchers used while the hypothesis to be tested is that the proportion of vouchers used in the English-speaking group is larger than the proportion in the non-English-speaking group. From the provided data, there are 37 people in each group; of the 1332 available vouchers for use in each group, the English-speaking participants used a total of 1269 while the non-English-speaking participants used a total of 1173. Using the Normal distribution, the test statistic for this test is z=1.121599 with a P-value of 0.131016. Using any of the standard significance levels, *our conclusion is that primary language does not have a statistically significant impact on spending.*

7. Does household size affect voucher usage for Healthy Options participants?

The students were particularly interested in whether larger families use more vouchers. To test this claim, many students categorized family size according to the groups 1-2, 3-4, 5-7. For each group, the mean number of vouchers used along with the standard deviation is shown below.

Household size	1-2	3-4	5-7
Mean	32.6	32.85714286	33.66666667
Standard deviation	6.976149845	5.140358005	4.651164729

Given the group sizes of 25, 28 and 21, respectively, it is unclear if there is a statistically significant difference in voucher used just from cursory inspection. To address this issue more fully, we run an ANOVA test, with confidence in the accuracy of the result since the data and groups adhere to the required assumptions well enough. The starting assumption is that there is no difference in the average number of vouchers used within each category of household size. The hypothesis we want to test is that there is a difference, with a suspicion that the difference lies with larger households. The F-statistic is 0.213310 with a P-value of 0.809207. This suggests that the differences in the data can be explained by natural variation. *Thus we do not have evidence to suggest that household size affects voucher usage.*