

Learning a Codebase

Derek Somerville

December 4, 2025

Contents

1 Repository	3
1.1 Glossary - Summary	3
2 Repository: 14	4
2.1 Box plot developer - 14 For packages touched for each period	4
2.2 Percentage touched reached - 14 For component touched for to reach 70.0 percentage known .	10
2.3 Percentage touched reached - 14 For component touched for to reach 80.0 percentage known .	13
2.4 Percentage touched reached - 14 For component touched for to reach 90.0 percentage known .	16
3 Repository: 21	19
3.1 Box plot developer - 21 For packages touched for each period	19
3.2 Percentage touched reached - 21 For component touched for to reach 70.0 percentage known .	25
3.3 Percentage touched reached - 21 For component touched for to reach 80.0 percentage known .	28
3.4 Percentage touched reached - 21 For component touched for to reach 90.0 percentage known .	31
4 Repository: 3	34
5 Repository: 5	34
6 Repository: 7	34
7 Repository: 10	34
8 Repository: 15	34
9 Repository: 18	34
10 Repository: 20	34
11 Repository: 26	34
12 Repository: 28	34
13 Repository: 33	34
14 Repository: 36	34
15 Repository: 37	34
16 Repository: 39	34
16.1 Smooth moving average - 14 For packages touched for each unit	34
16.2 Smooth moving average - 21 For packages touched for each unit	46
17 Repository pull request - Pull requests	58

1 Repository

1.1 Glossary - Summary

- The founder developer starts in the first six months of a project.
- Late joiner developers begin after six months.
- Transient developers have ten (10) or fewer commits.
- Moderate developers have more than ten (10) commits but fewer than 50 commits or commits for less than 250 days.
- Sustained developers make 50 or more commits and commit for 250 days or more.

2 Repository: 14

2.1 Box plot developer - 14 For packages touched for each period

A box plot of packages touched on average each period the number of commits to components touched.

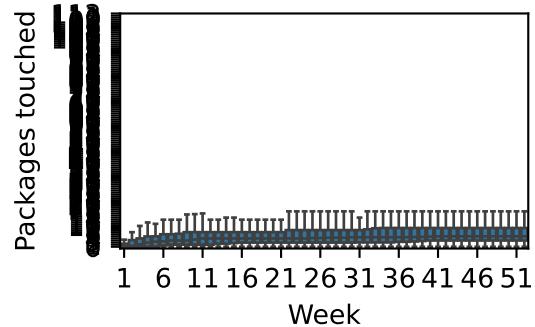
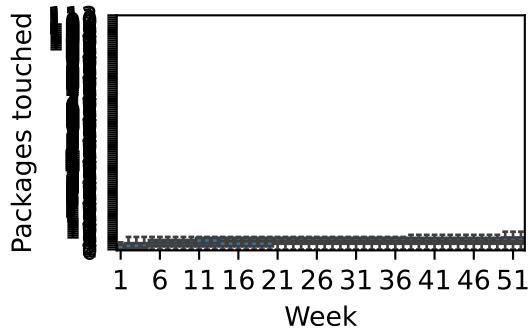
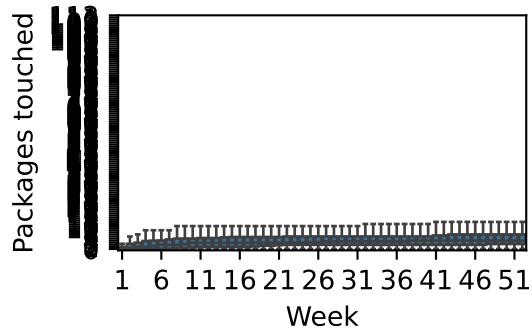


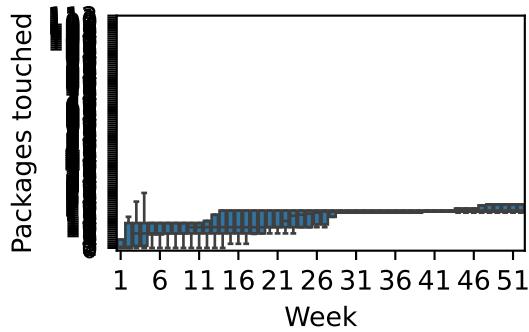
Figure 1: Repository: 14. A time series of the average (mean) total packages touched (y-axis) against the number of weeks (x-axis) for all developers (n=154).



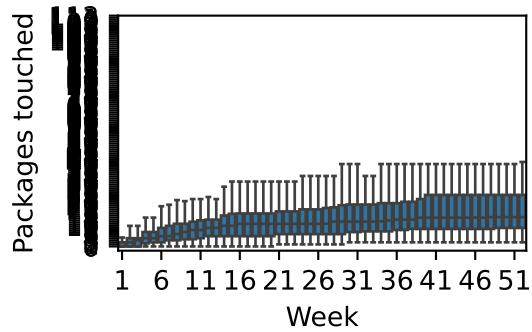
(a) Transient founder developers (n=8)



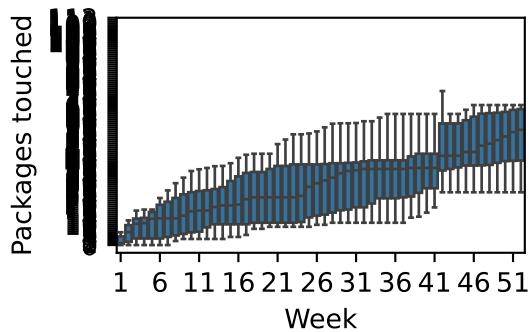
(b) Transient later joiner developers (n=119)



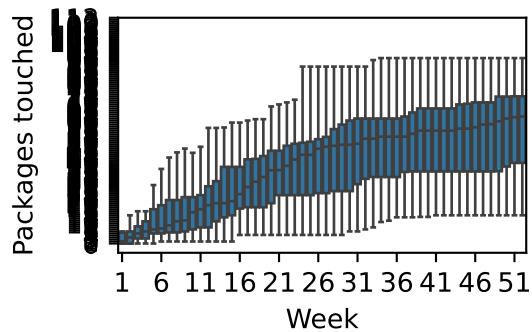
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 2: Repository: 14. A box plot of total packages touched on mean each month, with quartile shading.

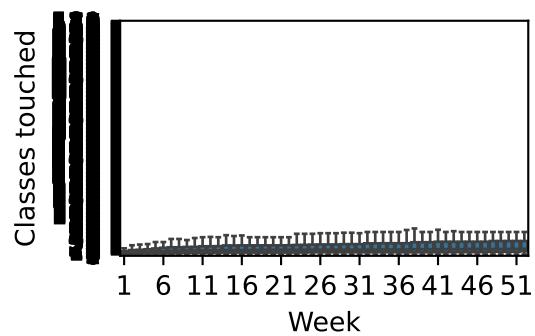
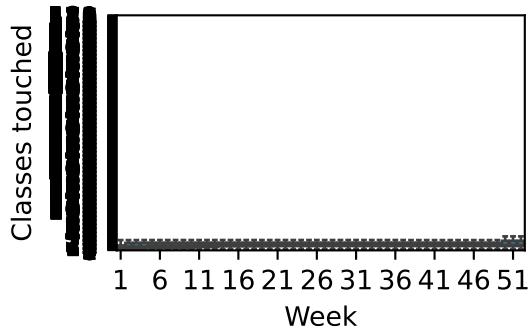
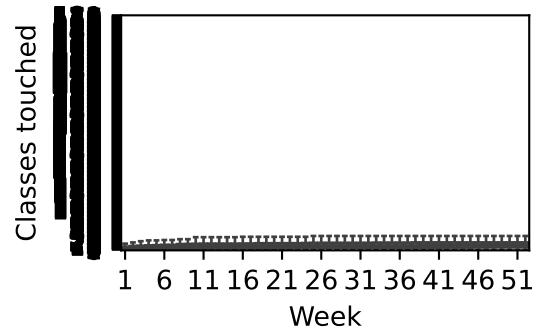


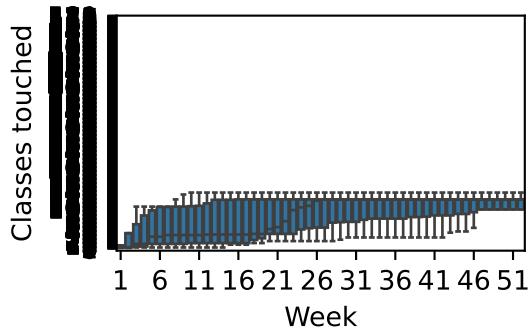
Figure 3: Repository: 14. A time series of the average (mean) total classes touched (y-axis) against the number of weeks (x-axis) for all developers (n=154).



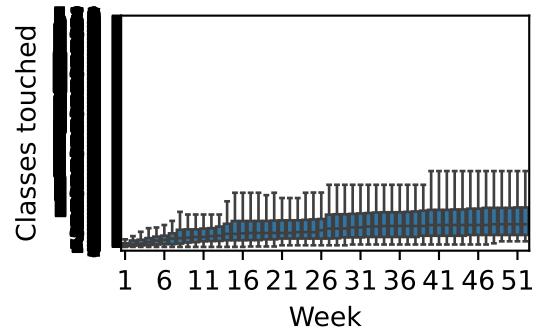
(a) Transient founder developers (n=8)



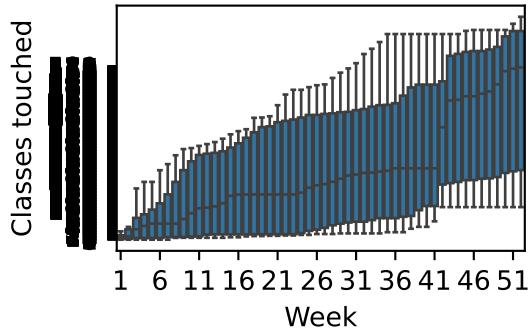
(b) Transient later joiner developers (n=119)



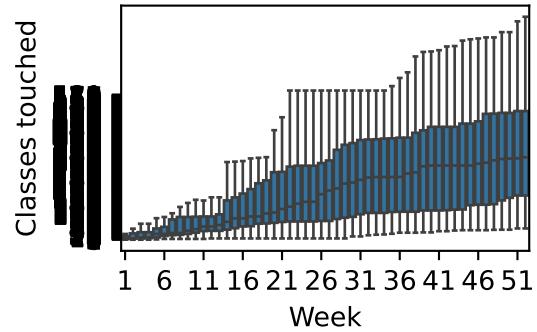
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 4: Repository: 14. A box plot of total classes touched on mean each month, with quartile shading.

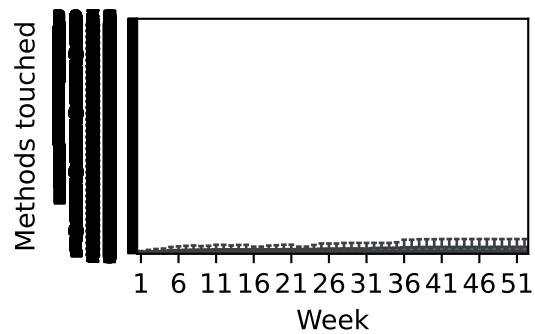


Figure 5: Repository: 14. A time series of the average (mean) total methods touched (y-axis) against the number of weeks (x-axis) for all developers (n=154).

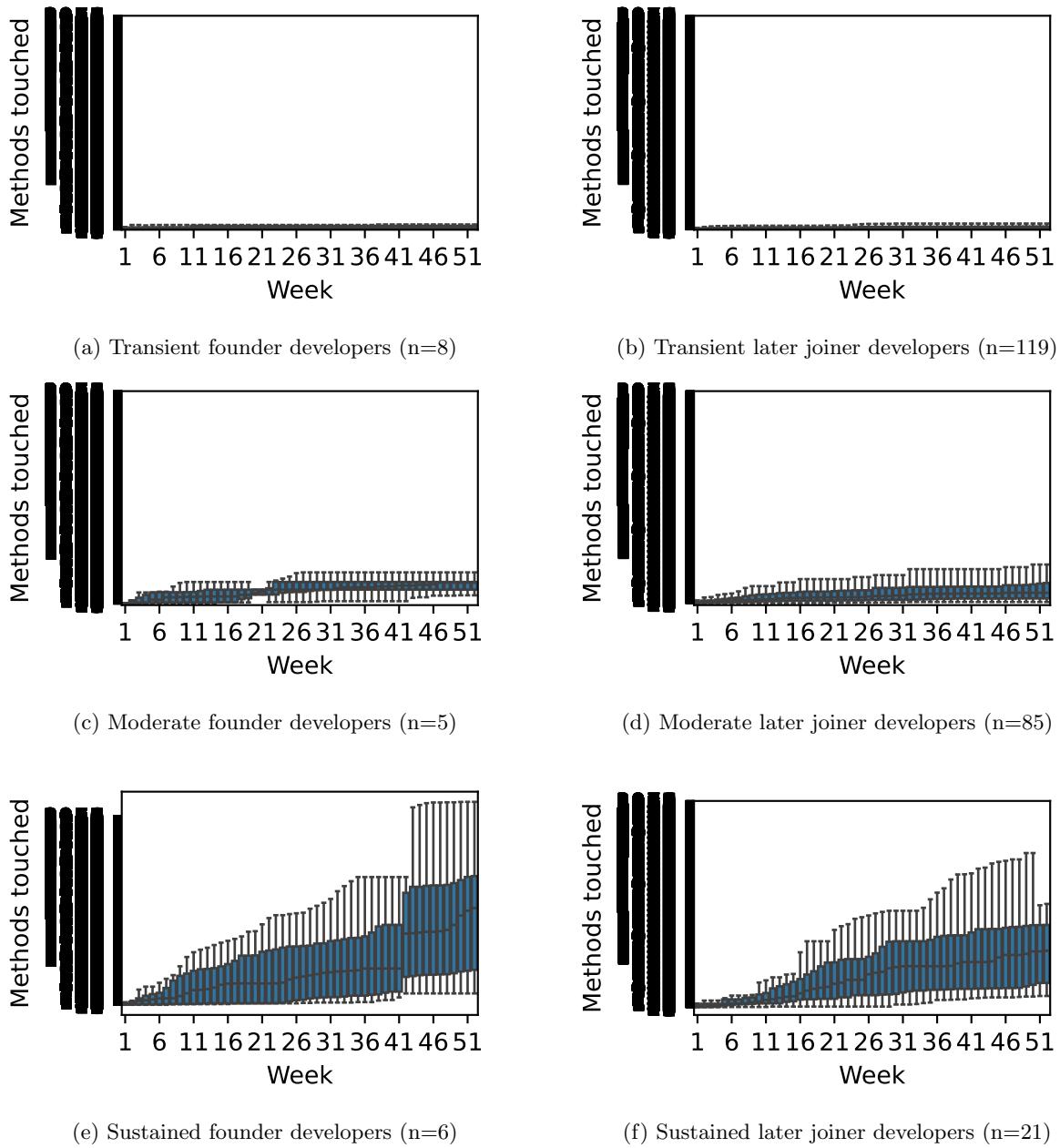
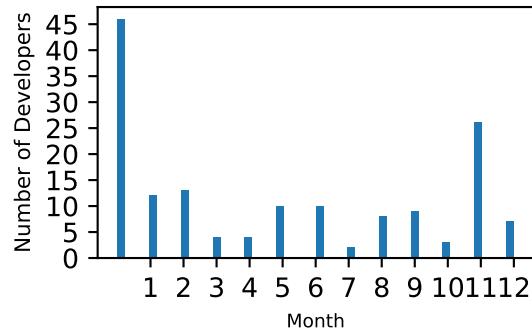


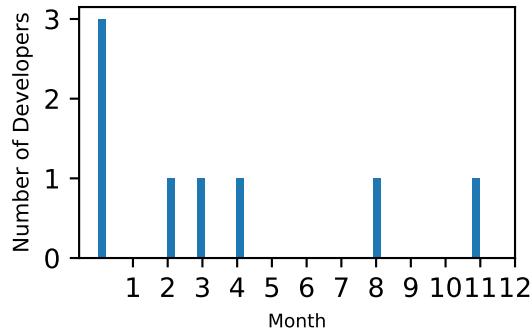
Figure 6: Repository: 14. A box plot of total methods touched on mean each month, with quartile shading.

2.2 Percentage touched reached - 14 For component touched for to reach 70.0 percentage known

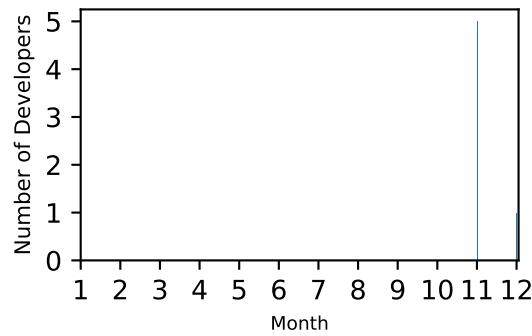
A histogram of number of developers that have touched a percentage of packages each period



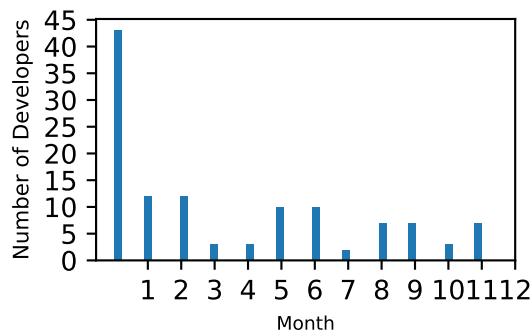
(a) All



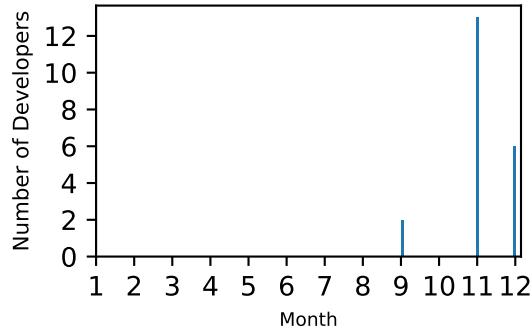
(b) transient founder



(c) sustained founder



(d) transient later joiner



(e) sustained later joiner

Figure 7: A histogram of number of developers that have touched a percentage (70.0%) of packages each month of a year.

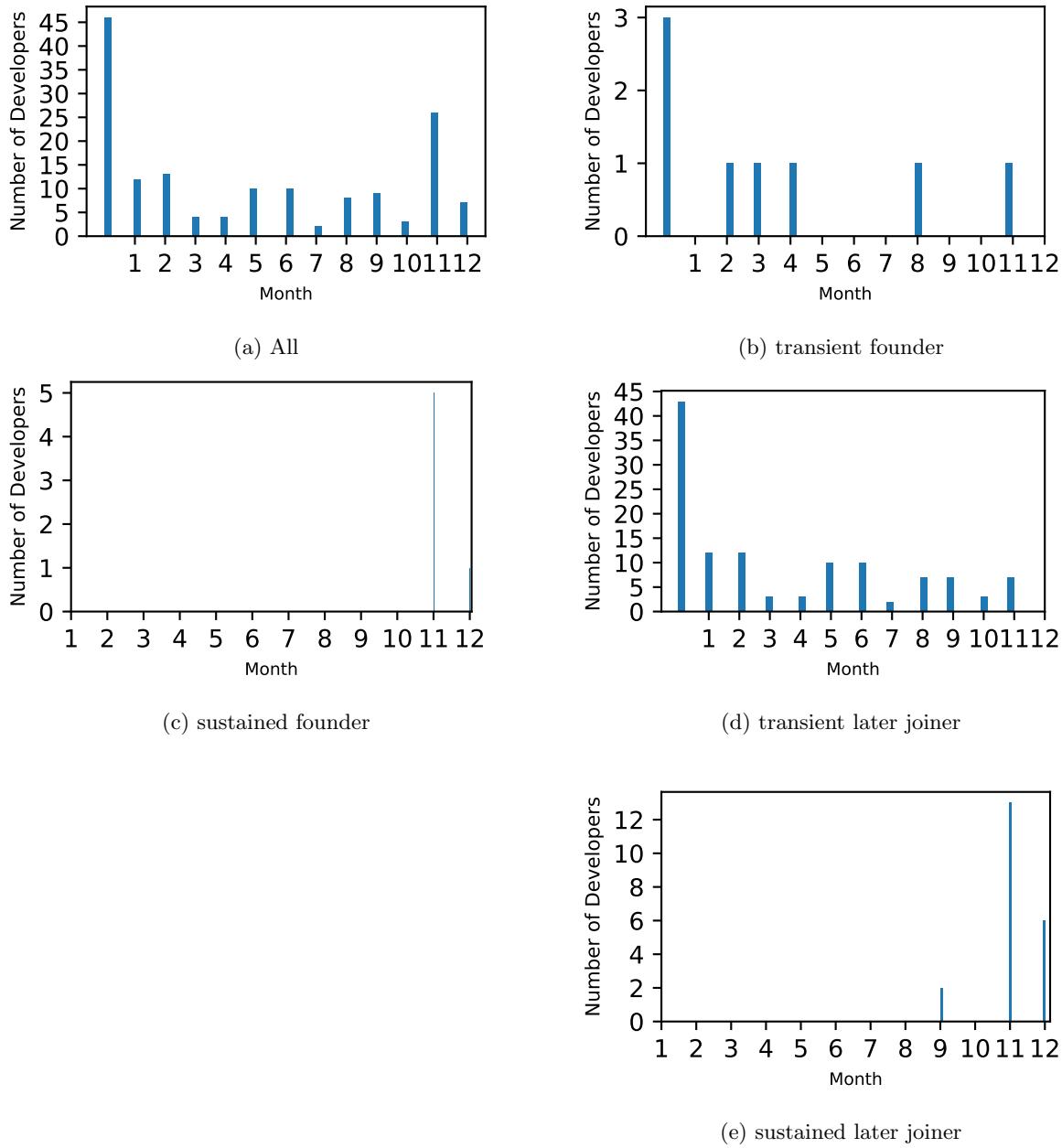
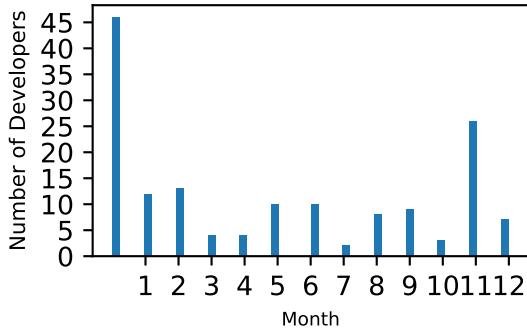
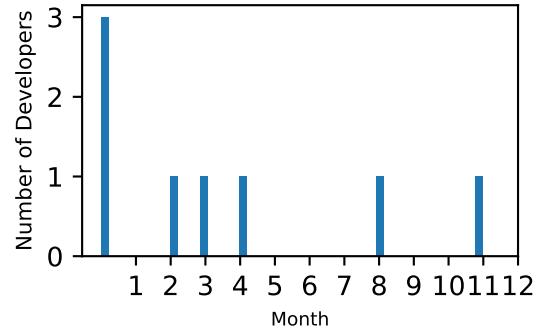


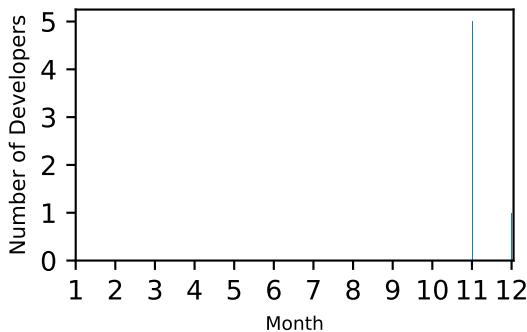
Figure 8: A histogram of number of developers that have touched a percentage (70.0%) of classes each month of a year.



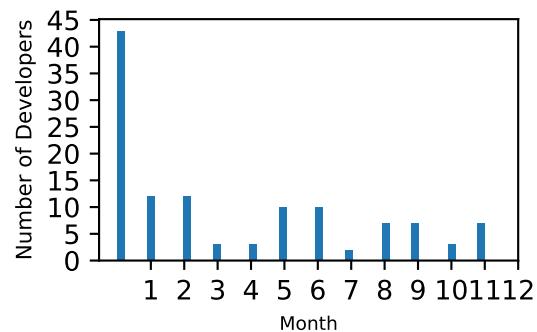
(a) All



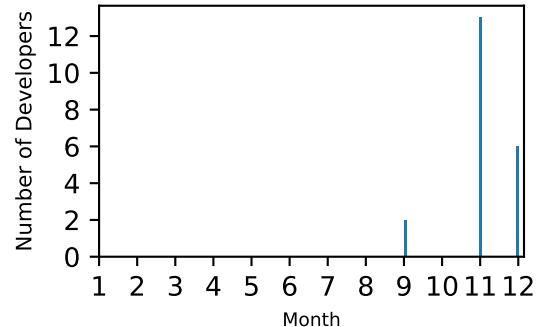
(b) transient founder



(c) sustained founder



(d) transient later joiner

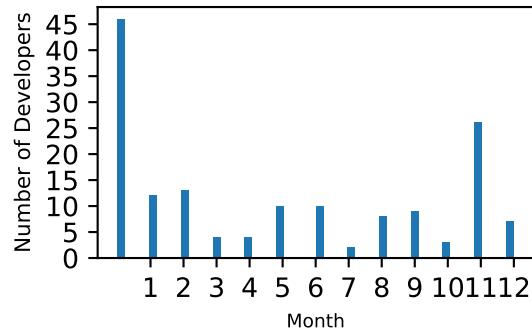


(e) sustained later joiner

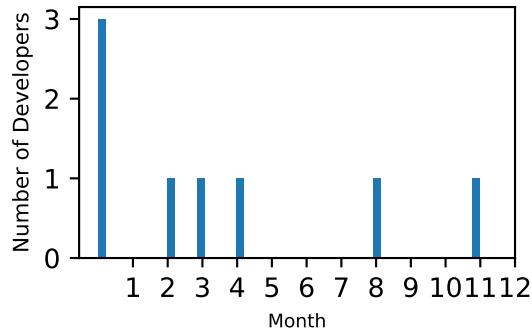
Figure 9: A histogram of number of developers that have touched a percentage (70.0%) of methods each month of a year.

2.3 Percentage touched reached - 14 For component touched for to reach 80.0 percentage known

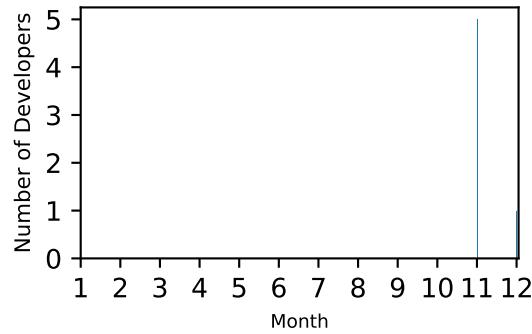
A histogram of number of developers that have touched a percentage of packages each period



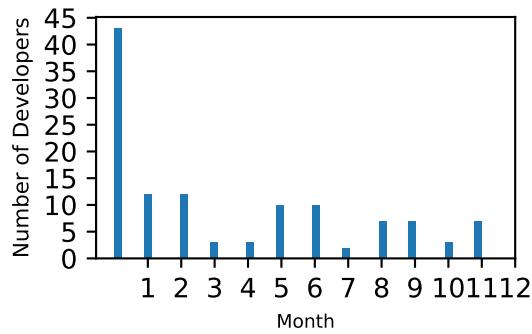
(a) All



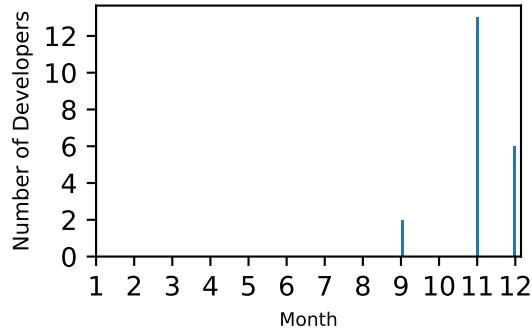
(b) transient founder



(c) sustained founder

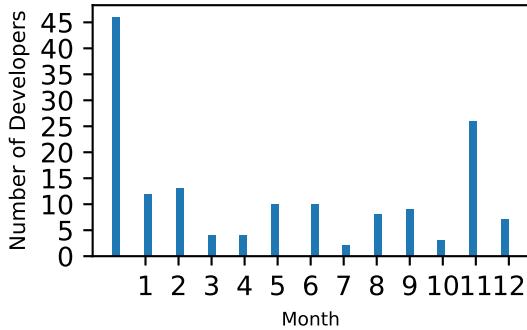


(d) transient later joiner

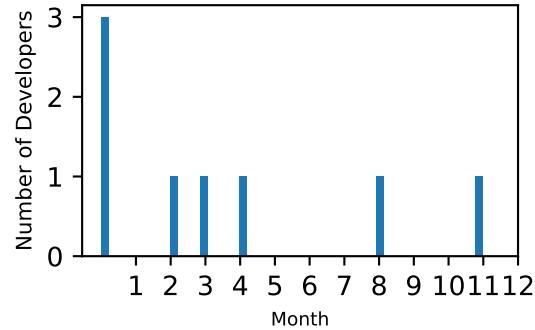


(e) sustained later joiner

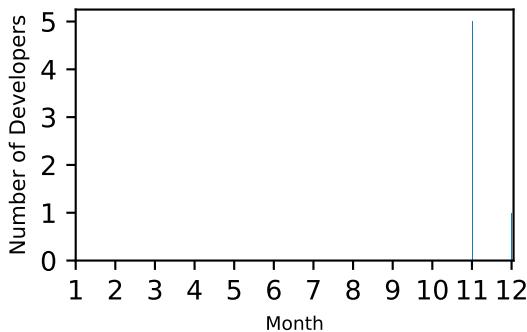
Figure 10: A histogram of number of developers that have touched a percentage (80.0%) of packages each month of a year.



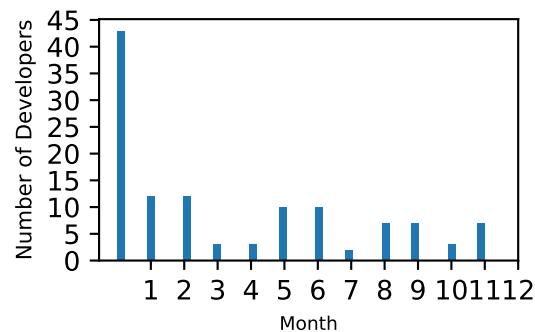
(a) All



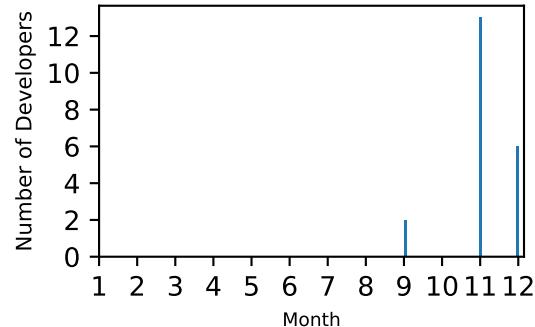
(b) transient founder



(c) sustained founder

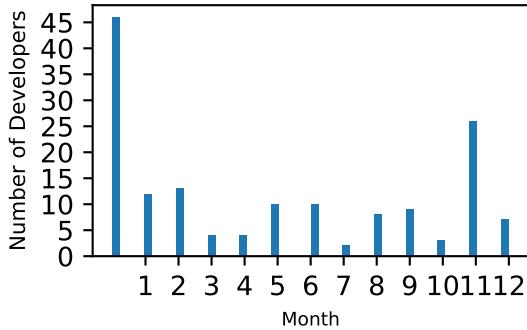


(d) transient later joiner

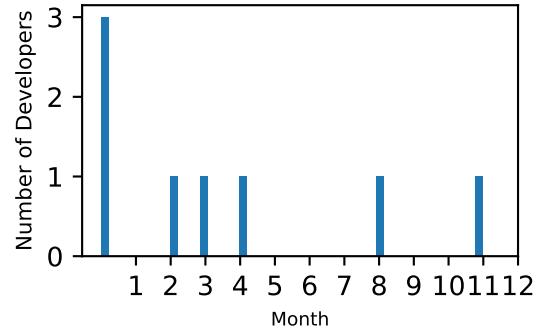


(e) sustained later joiner

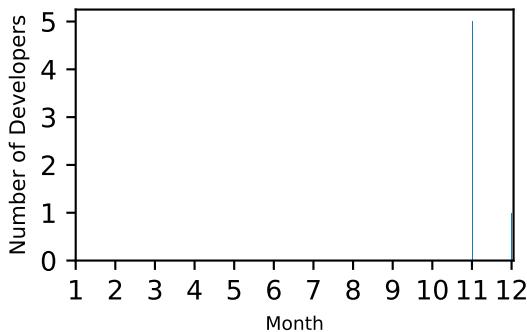
Figure 11: A histogram of number of developers that have touched a percentage (80.0%) of classes each month of a year.



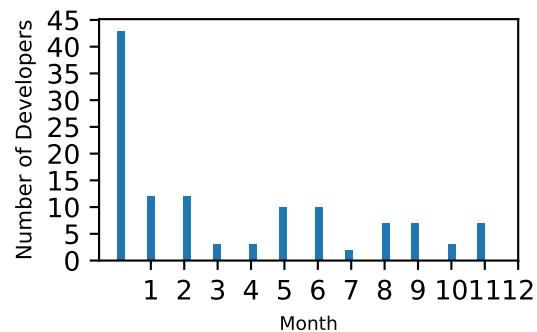
(a) All



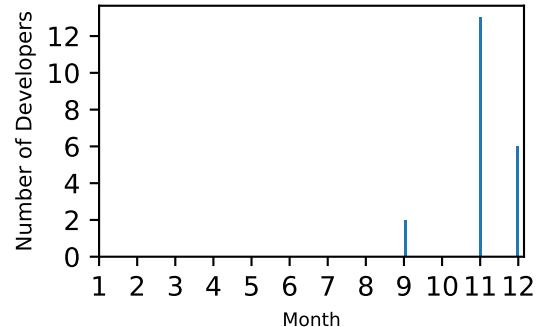
(b) transient founder



(c) sustained founder



(d) transient later joiner

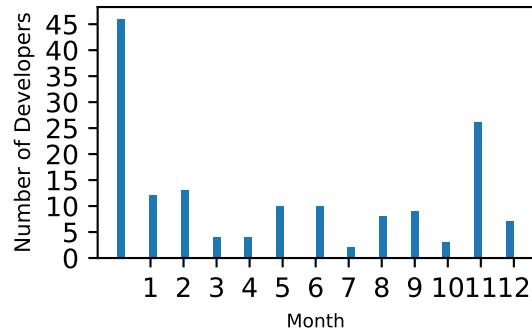


(e) sustained later joiner

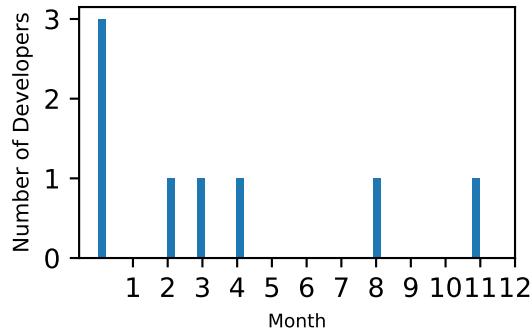
Figure 12: A histogram of number of developers that have touched a percentage (80.0%) of methods each month of a year.

2.4 Percentage touched reached - 14 For component touched for to reach 90.0 percentage known

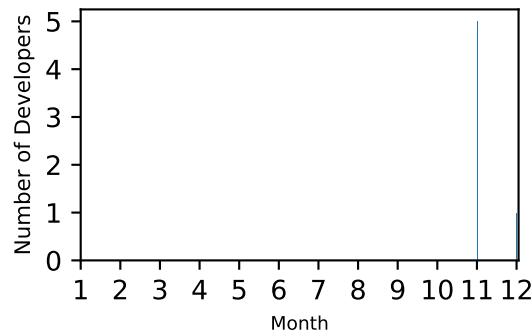
A histogram of number of developers that have touched a percentage of packages each period



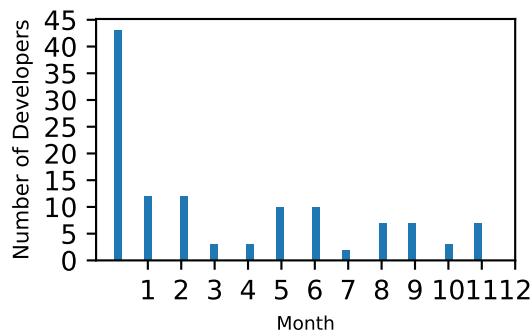
(a) All



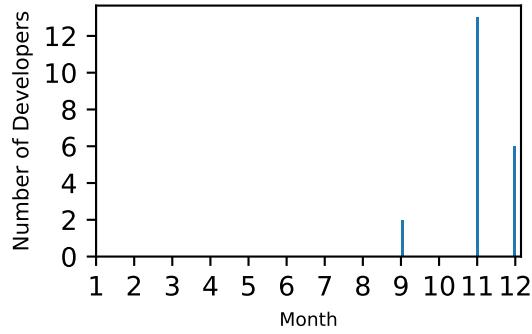
(b) transient founder



(c) sustained founder

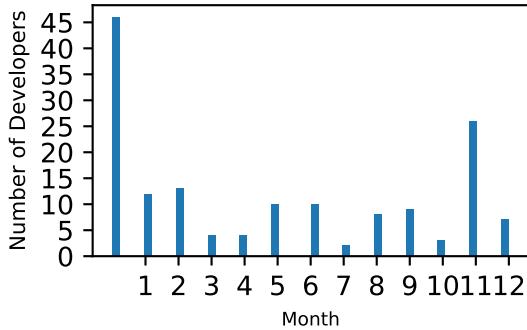


(d) transient later joiner

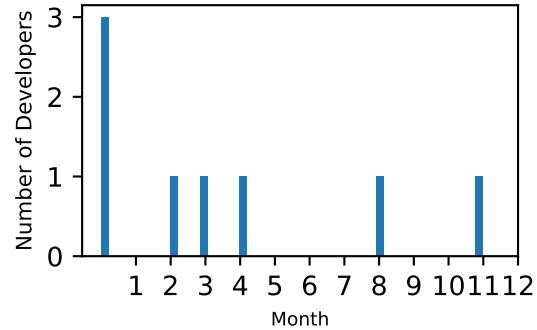


(e) sustained later joiner

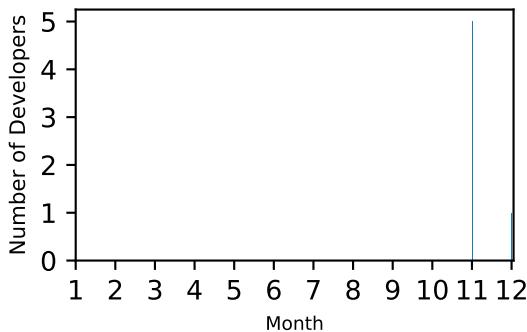
Figure 13: A histogram of number of developers that have touched a percentage (90.0%) of packages each month of a year.



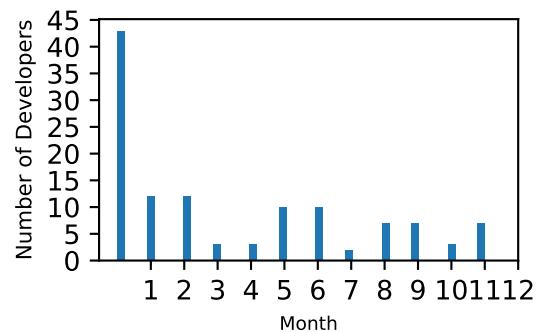
(a) All



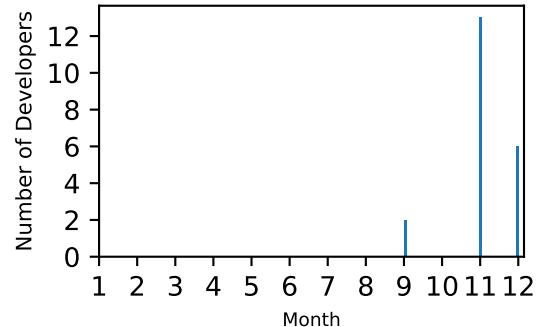
(b) transient founder



(c) sustained founder

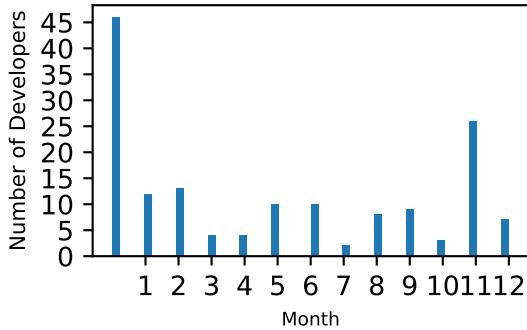


(d) transient later joiner

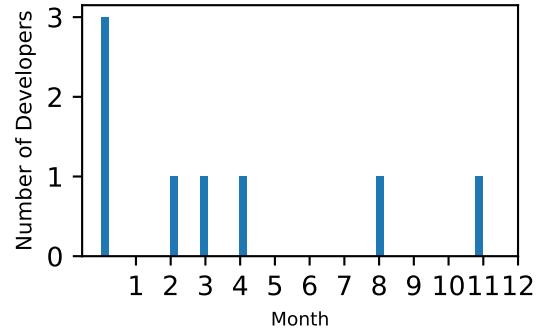


(e) sustained later joiner

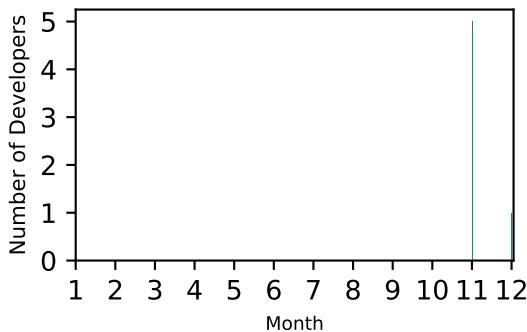
Figure 14: A histogram of number of developers that have touched a percentage (90.0%) of classes each month of a year.



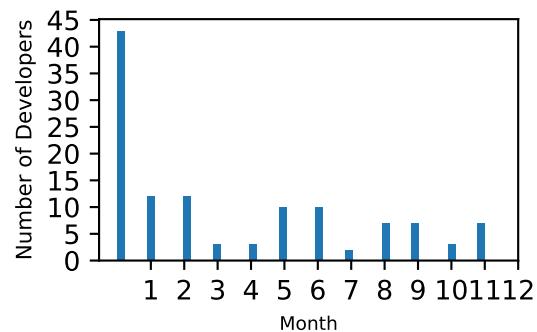
(a) All



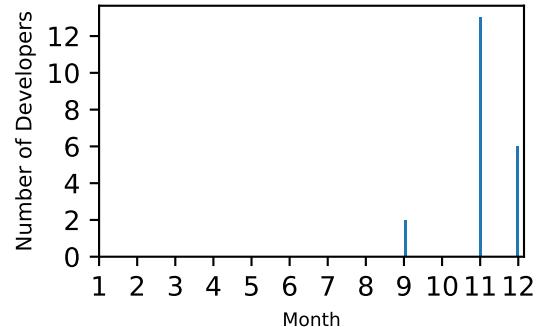
(b) transient founder



(c) sustained founder



(d) transient later joiner



(e) sustained later joiner

Figure 15: A histogram of number of developers that have touched a percentage (90.0%) of methods each month of a year.

3 Repository: 21

3.1 Box plot developer - 21 For packages touched for each period

A box plot of packages touched on average each period the number of commits to components touched.

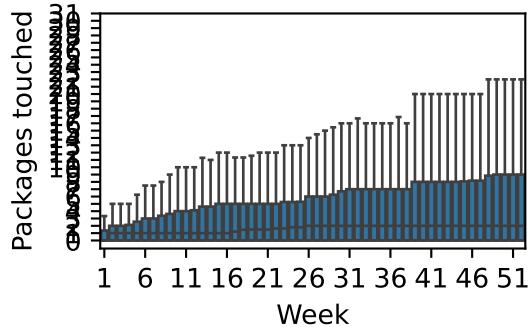
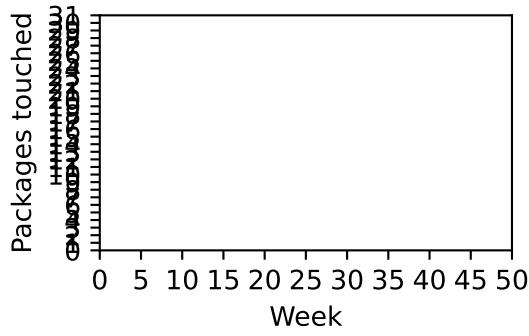
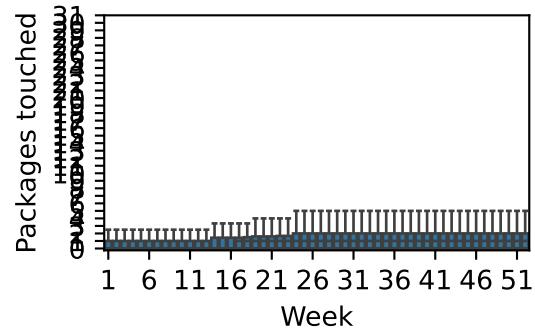


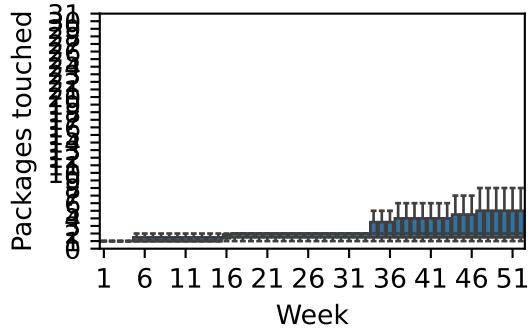
Figure 16: Repository: 21. A time series of the average (mean) total packages touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).



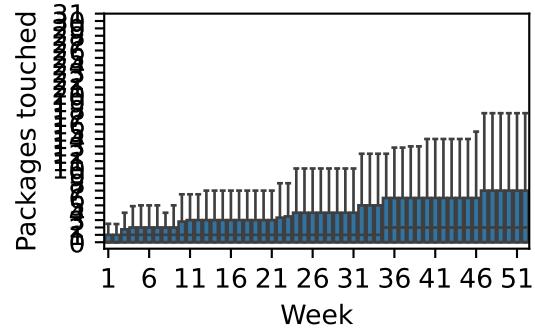
(a) Transient founder developers (n=0)



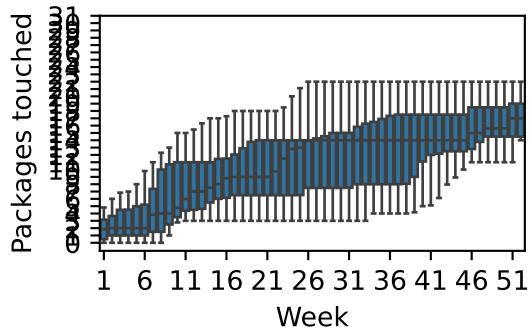
(b) Transient later joiner developers (n=113)



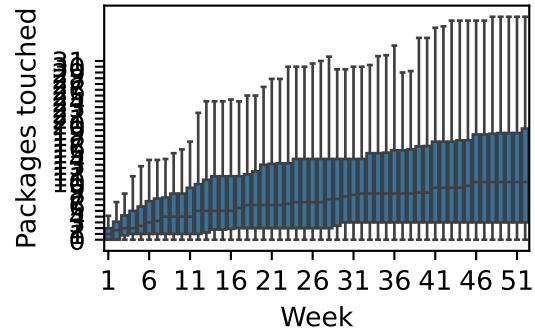
(c) Moderate founder developers (n=3)



(d) Moderate later joiner developers (n=93)



(e) Sustained founder developers (n=7)



(f) Sustained later joiner developers (n=84)

Figure 17: Repository: 21. A box plot of total packages touched on mean each month, with quartile shading.

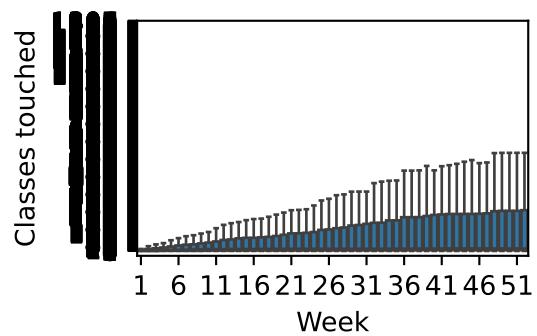


Figure 18: Repository: 21. A time series of the average (mean) total classes touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).

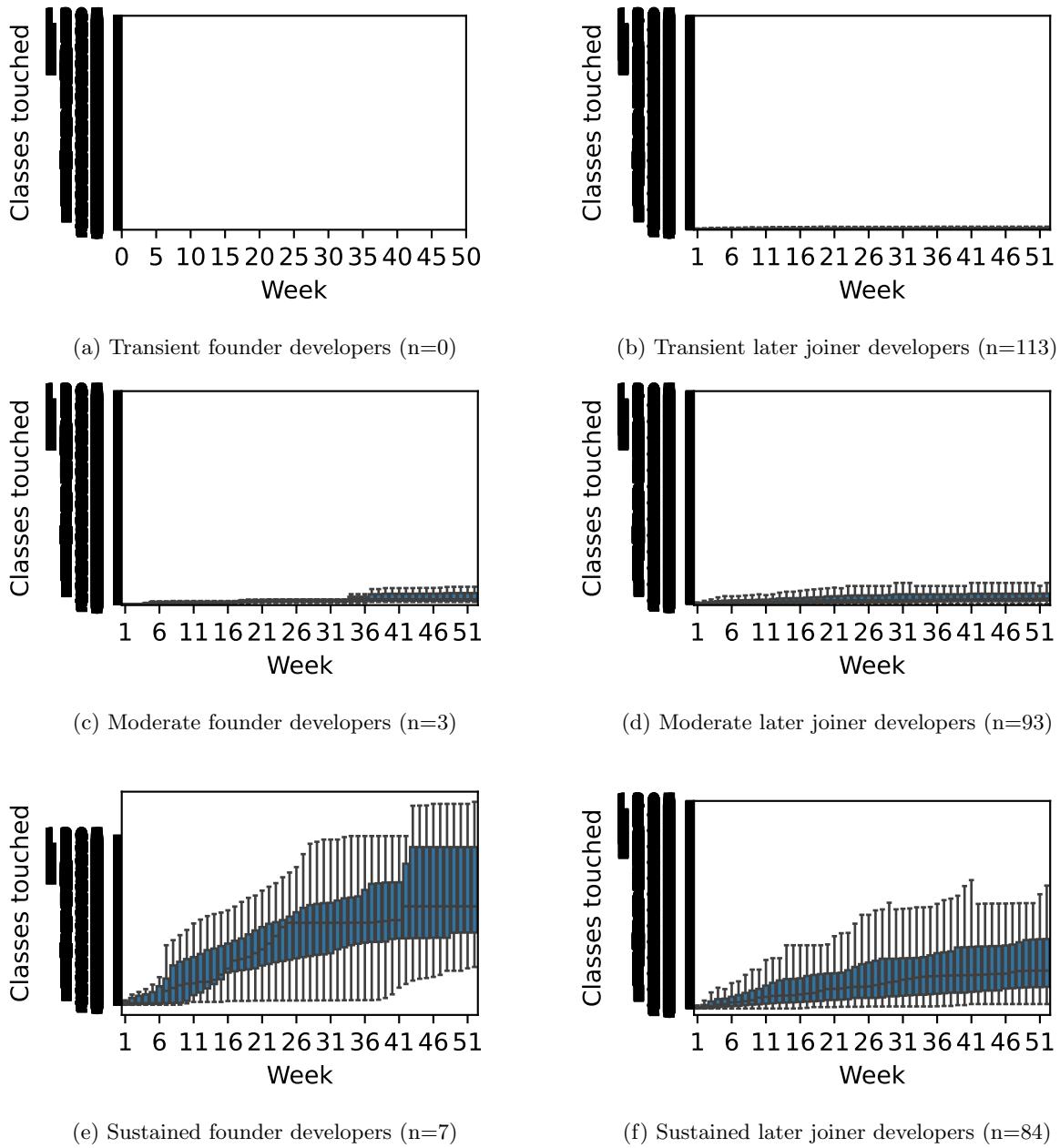


Figure 19: Repository: 21. A box plot of total classes touched on mean each month, with quartile shading.

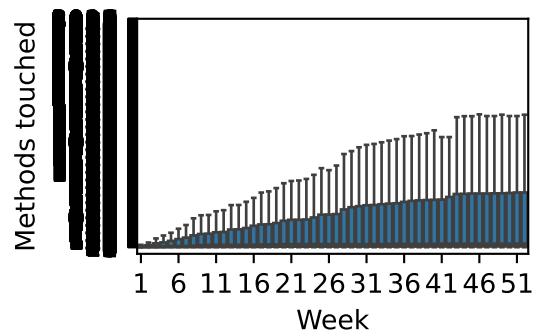


Figure 20: Repository: 21. A time series of the average (mean) total methods touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).

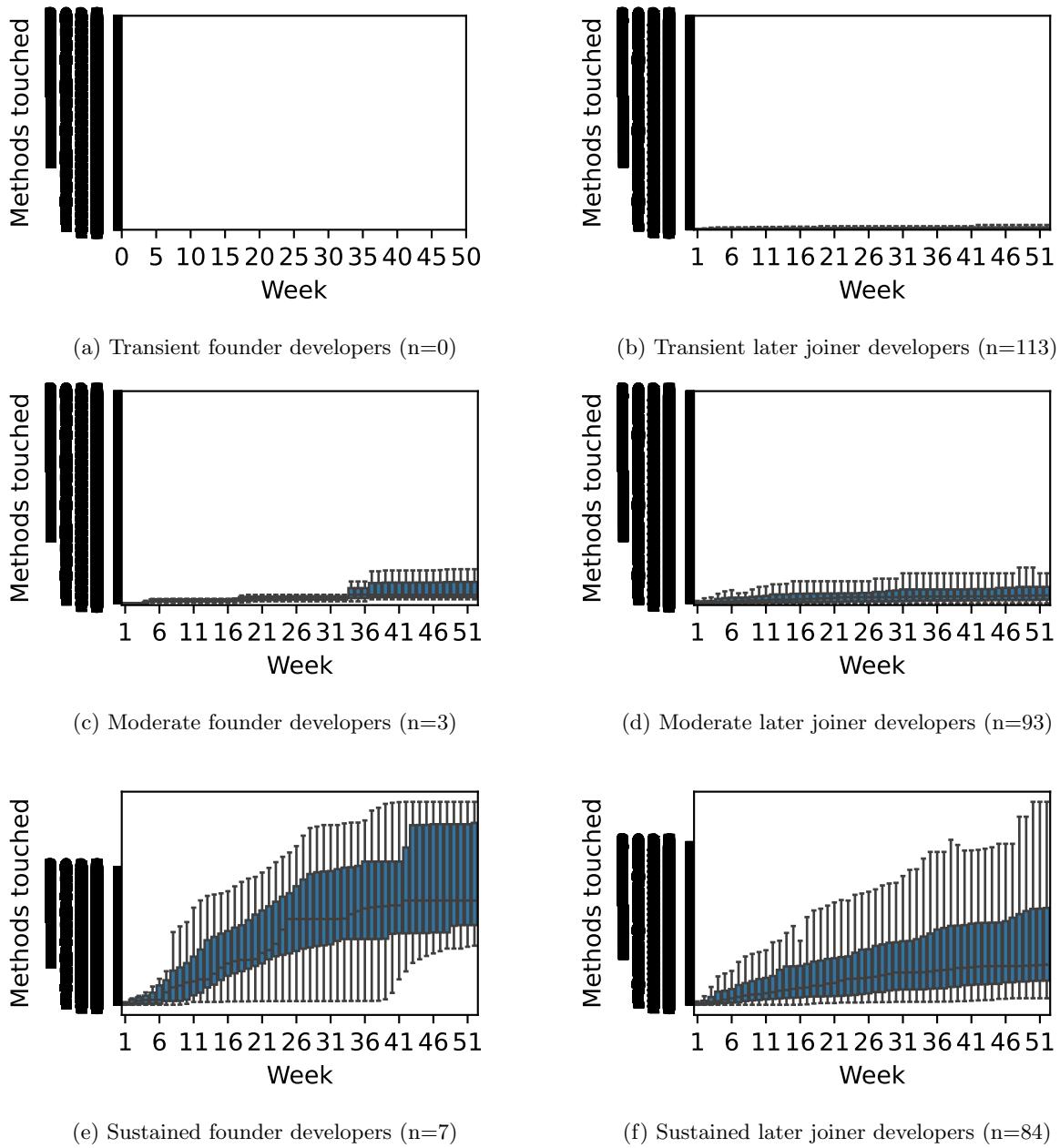
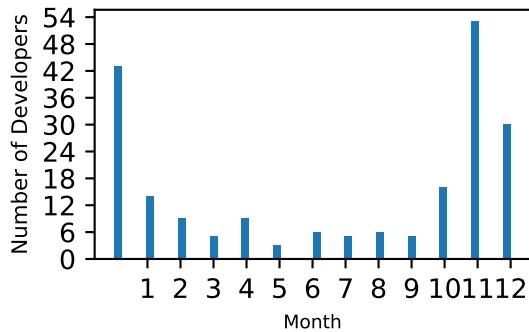


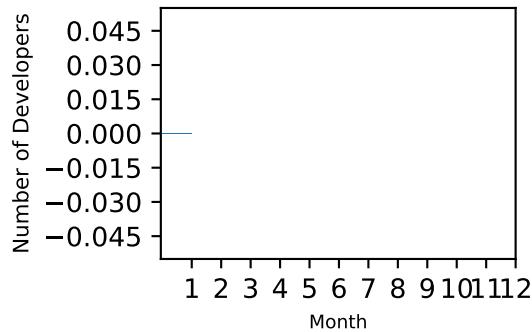
Figure 21: Repository: 21. A box plot of total methods touched on mean each month, with quartile shading.

3.2 Percentage touched reached - 21 For component touched for to reach 70.0 percentage known

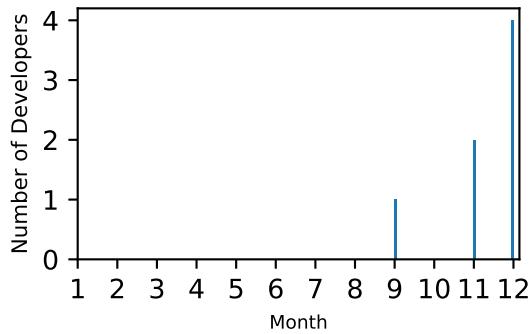
A histogram of number of developers that have touched a percentage of packages each period



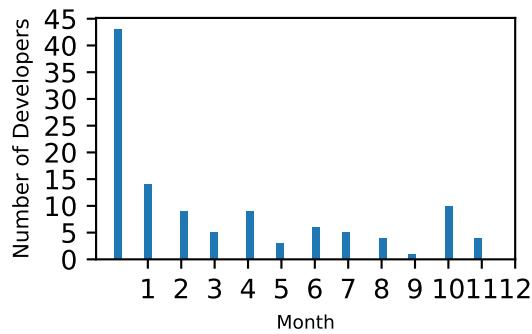
(a) All



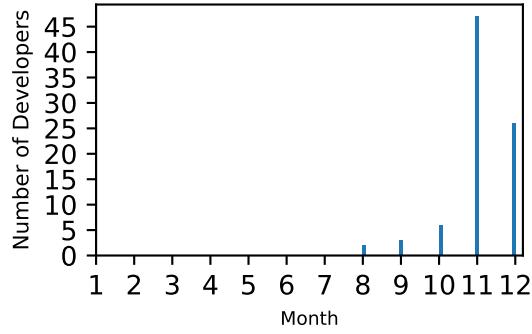
(b) transient founder



(c) sustained founder

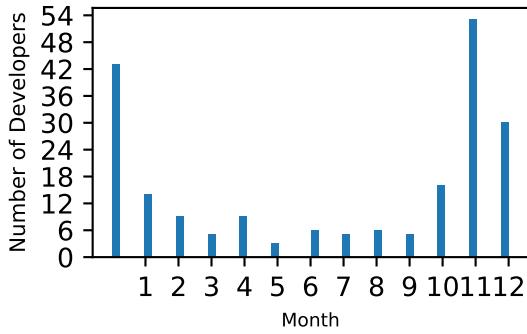


(d) transient later joiner

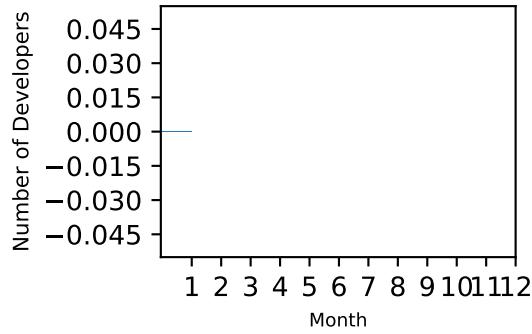


(e) sustained later joiner

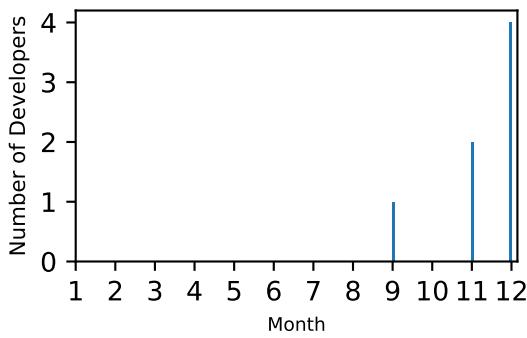
Figure 22: A histogram of number of developers that have touched a percentage (70.0%) of packages each month of a year.



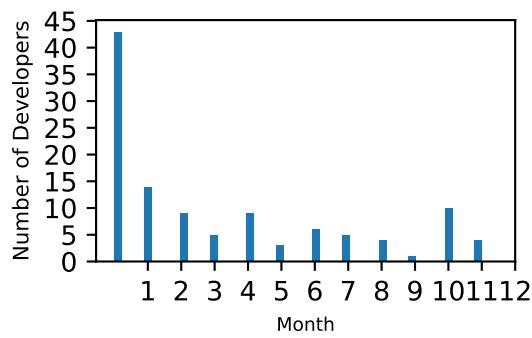
(a) All



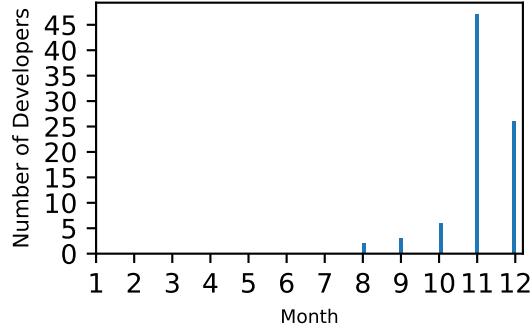
(b) transient founder



(c) sustained founder

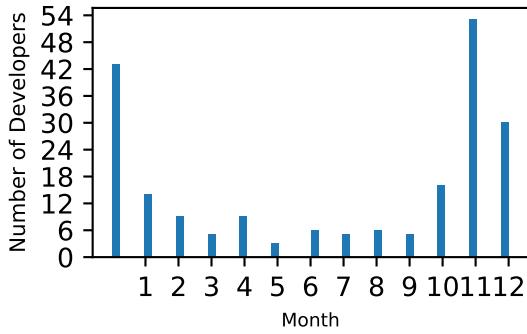


(d) transient later joiner

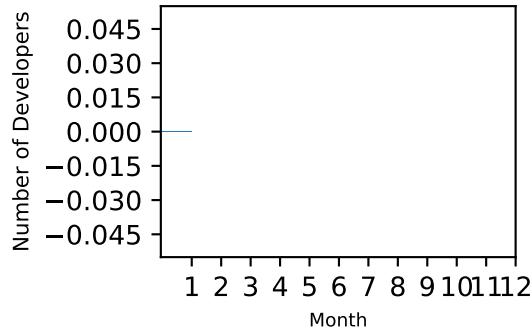


(e) sustained later joiner

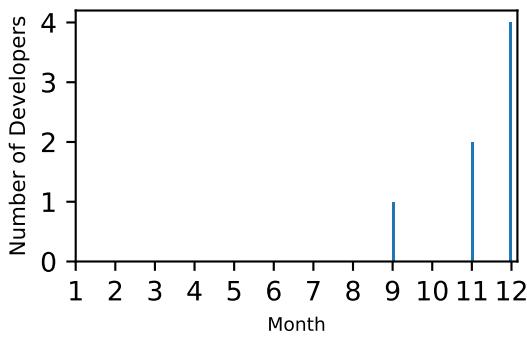
Figure 23: A histogram of number of developers that have touched a percentage (70.0%) of classes each month of a year.



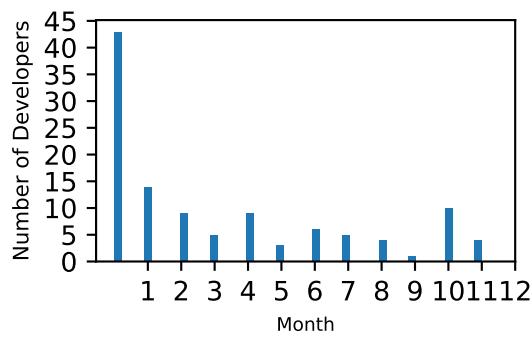
(a) All



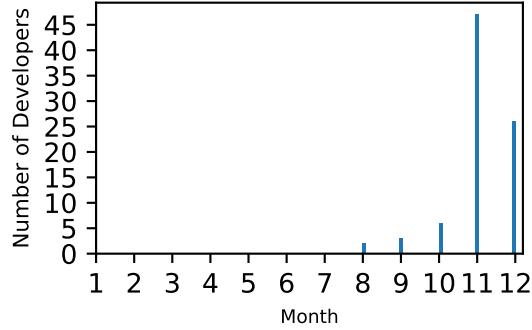
(b) transient founder



(c) sustained founder



(d) transient later joiner

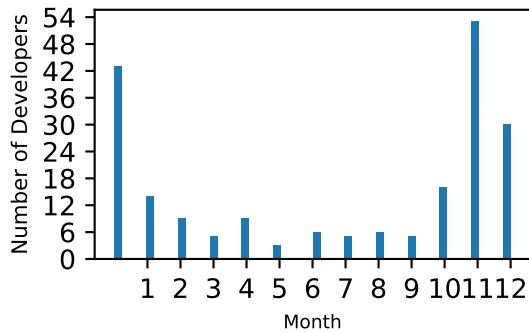


(e) sustained later joiner

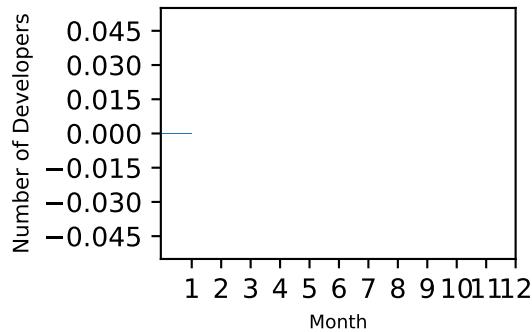
Figure 24: A histogram of number of developers that have touched a percentage (70.0%) of methods each month of a year.

3.3 Percentage touched reached - 21 For component touched for to reach 80.0 percentage known

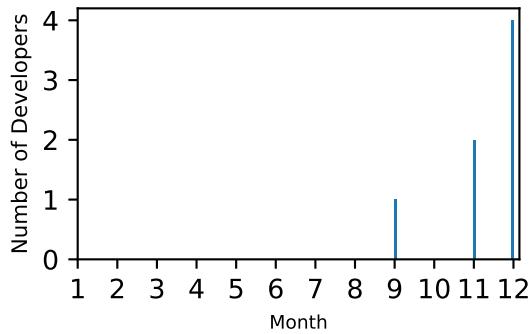
A histogram of number of developers that have touched a percentage of packages each period



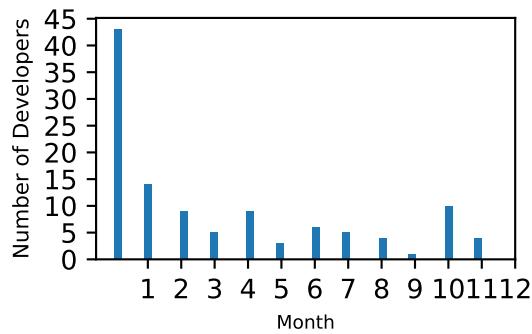
(a) All



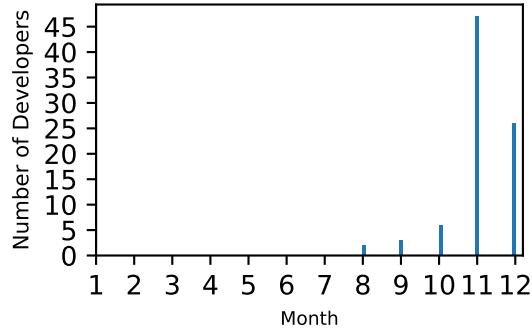
(b) transient founder



(c) sustained founder

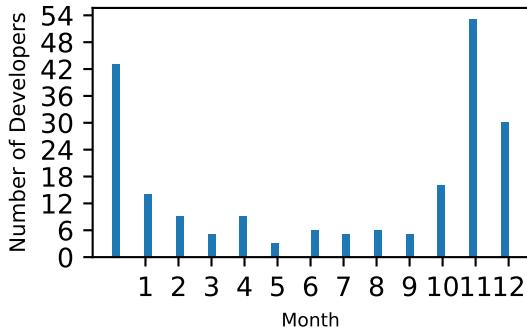


(d) transient later joiner

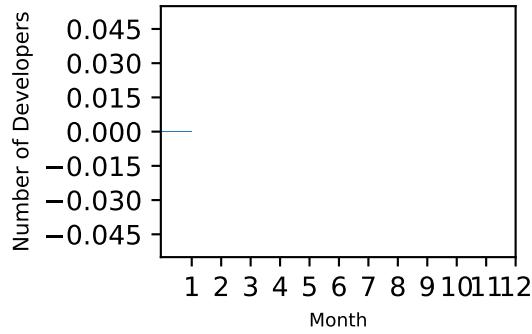


(e) sustained later joiner

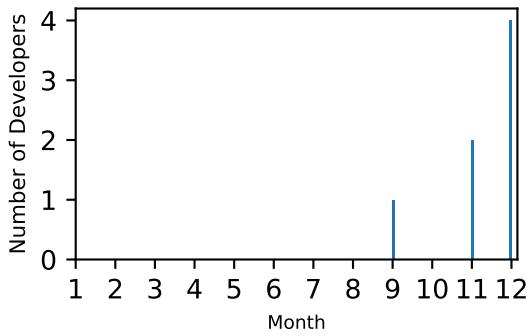
Figure 25: A histogram of number of developers that have touched a percentage (80.0%) of packages each month of a year.



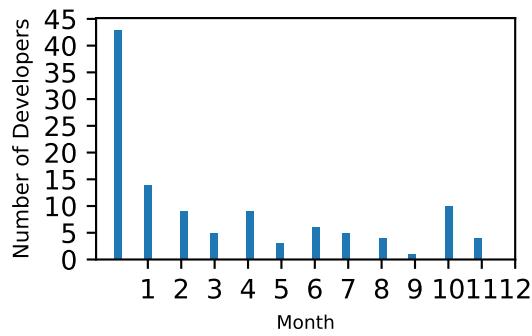
(a) All



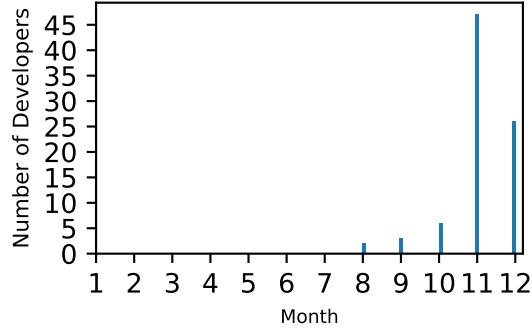
(b) transient founder



(c) sustained founder

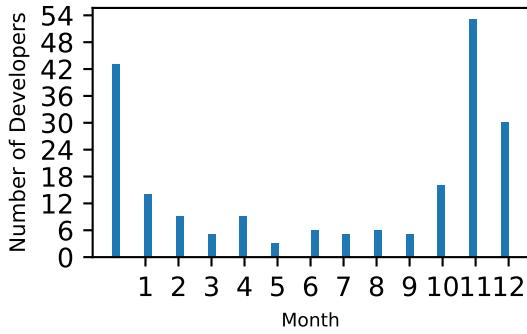


(d) transient later joiner

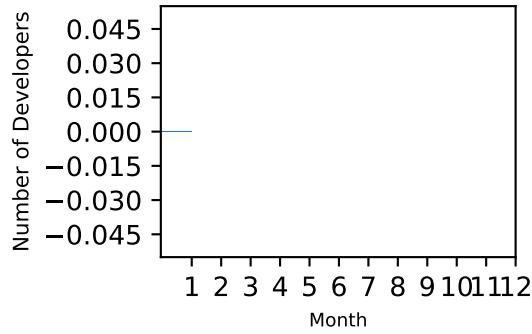


(e) sustained later joiner

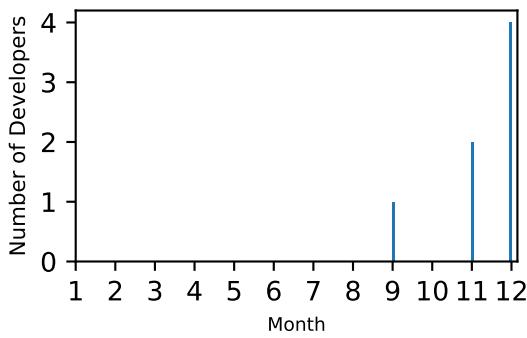
Figure 26: A histogram of number of developers that have touched a percentage (80.0%) of classes each month of a year.



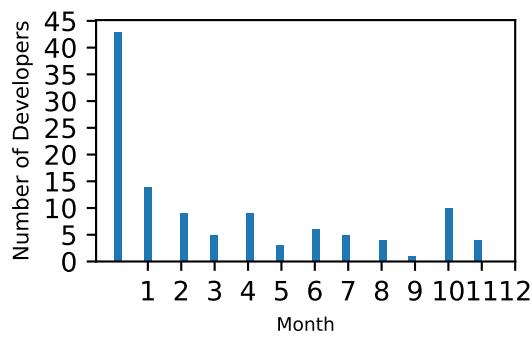
(a) All



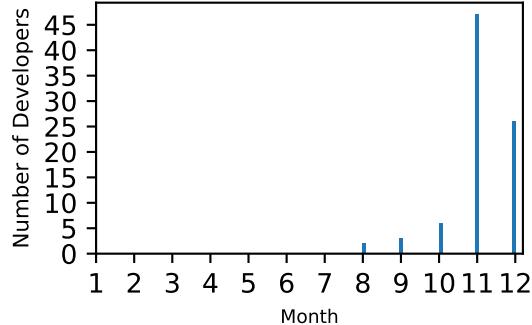
(b) transient founder



(c) sustained founder



(d) transient later joiner

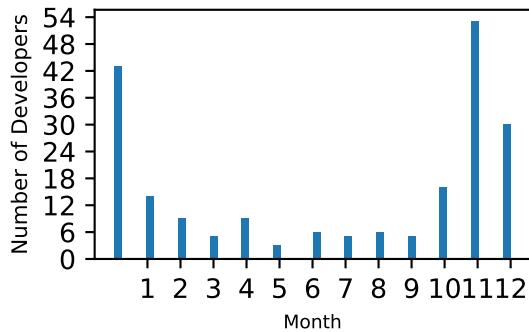


(e) sustained later joiner

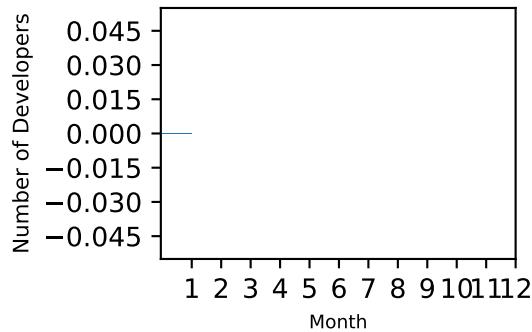
Figure 27: A histogram of number of developers that have touched a percentage (80.0%) of methods each month of a year.

3.4 Percentage touched reached - 21 For component touched for to reach 90.0 percentage known

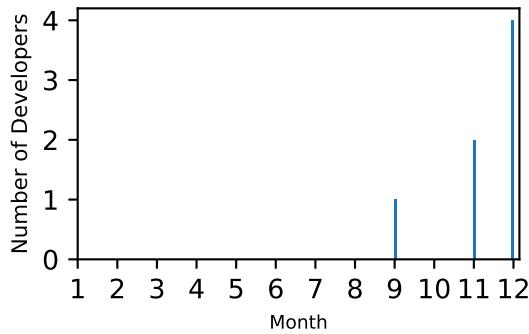
A histogram of number of developers that have touched a percentage of packages each period



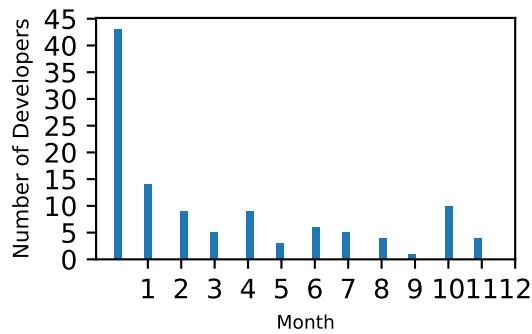
(a) All



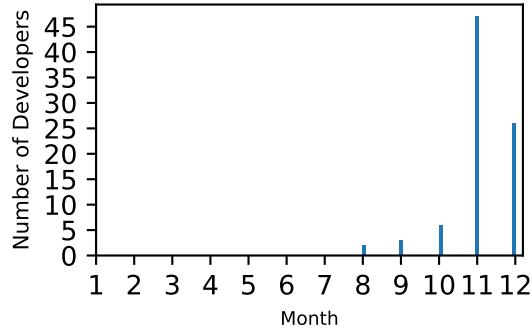
(b) transient founder



(c) sustained founder

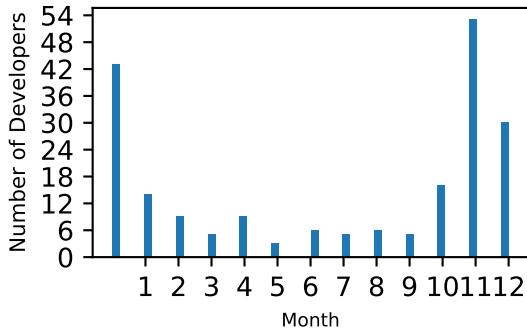


(d) transient later joiner

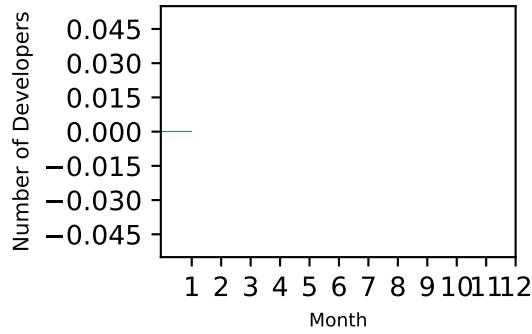


(e) sustained later joiner

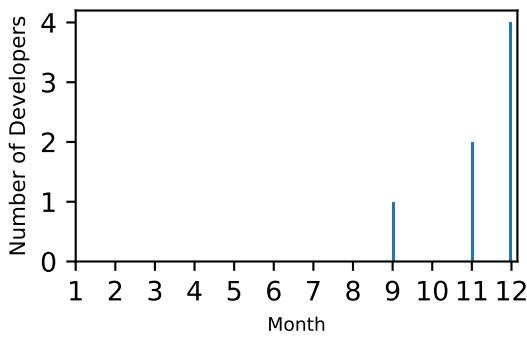
Figure 28: A histogram of number of developers that have touched a percentage (90.0%) of packages each month of a year.



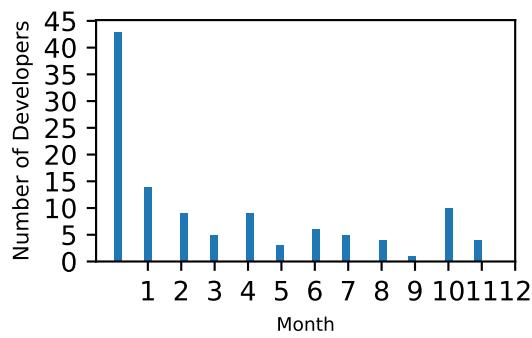
(a) All



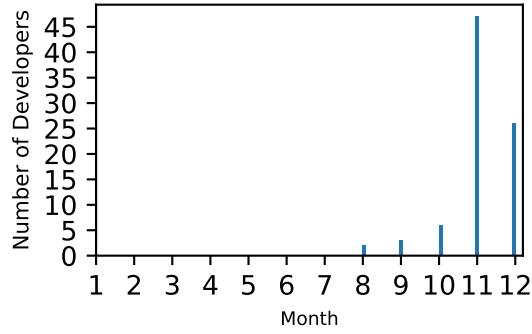
(b) transient founder



(c) sustained founder

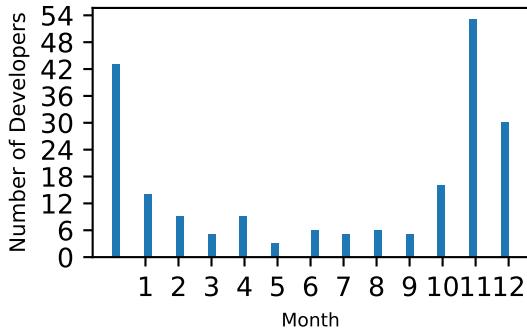


(d) transient later joiner

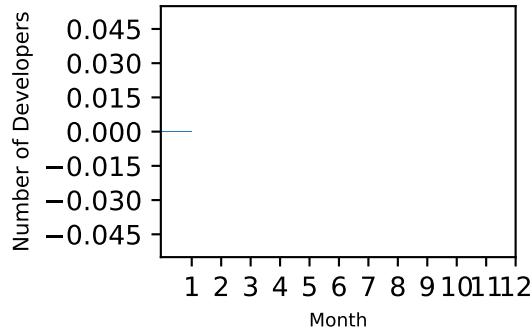


(e) sustained later joiner

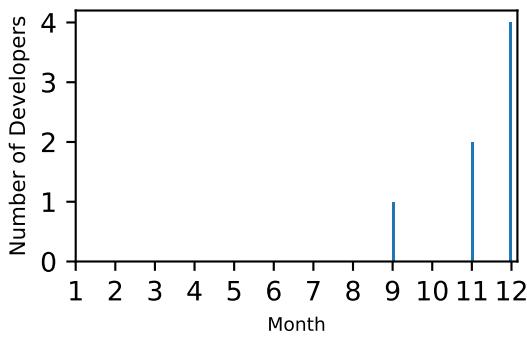
Figure 29: A histogram of number of developers that have touched a percentage (90.0%) of classes each month of a year.



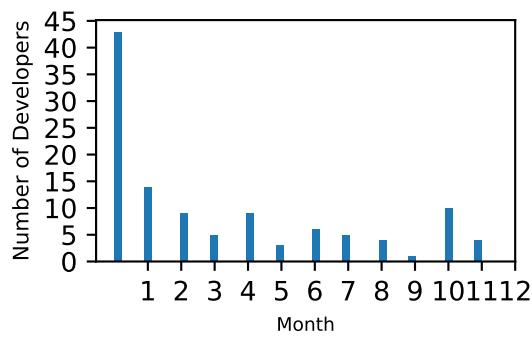
(a) All



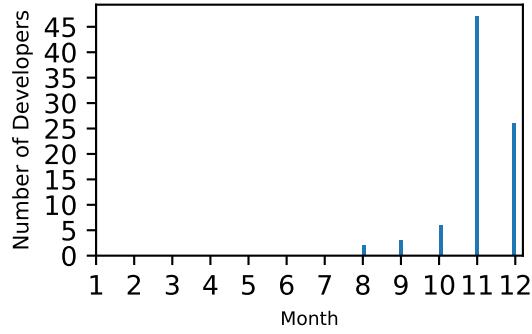
(b) transient founder



(c) sustained founder



(d) transient later joiner



(e) sustained later joiner

Figure 30: A histogram of number of developers that have touched a percentage (90.0%) of methods each month of a year.

- 4 Repository: 3**
- 5 Repository: 5**
- 6 Repository: 7**
- 7 Repository: 10**
- 8 Repository: 15**
- 9 Repository: 18**
- 10 Repository: 20**
- 11 Repository: 26**
- 12 Repository: 28**
- 13 Repository: 33**
- 14 Repository: 36**
- 15 Repository: 37**
- 16 Repository: 39**

16.1 Smooth moving average - 14 For packages touched for each unit

A smooth moving average packages touched on average each period the number of commits to components touched.

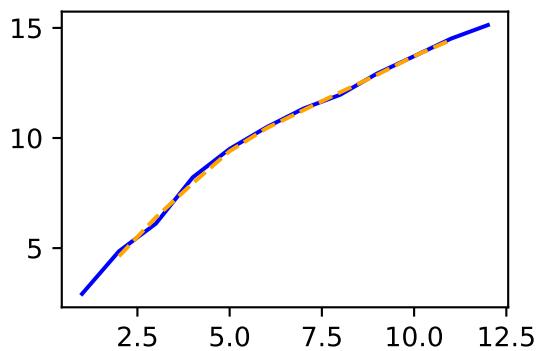
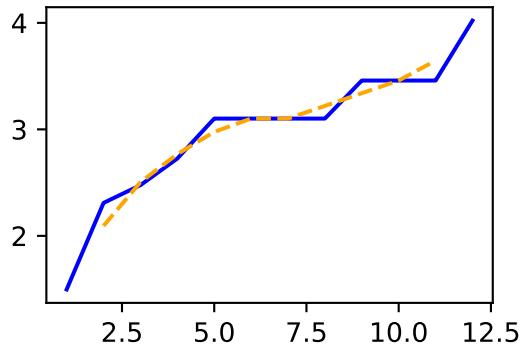
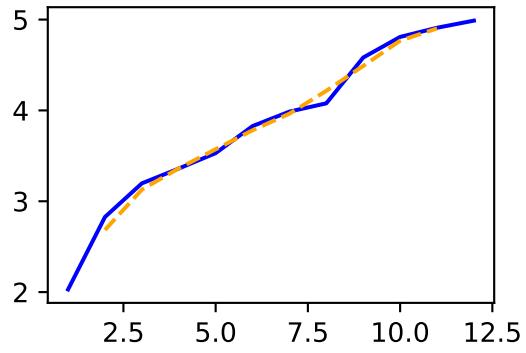


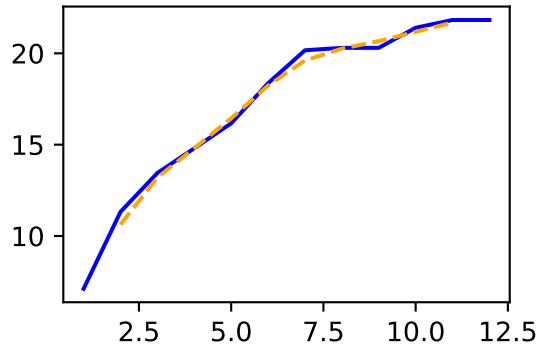
Figure 31: Repository: 14. A time series of the average (mean) total packages touched (y-axis) against the number of months (x-axis) for all developers (n=154).



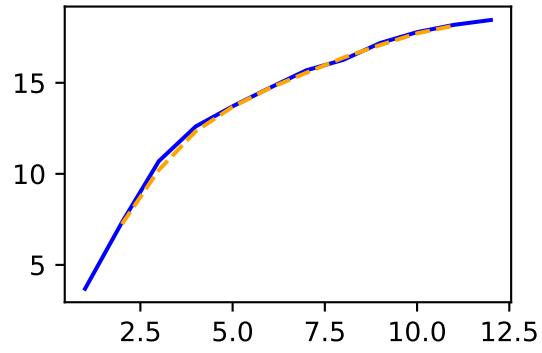
(a) Transient founder developers (n=8)



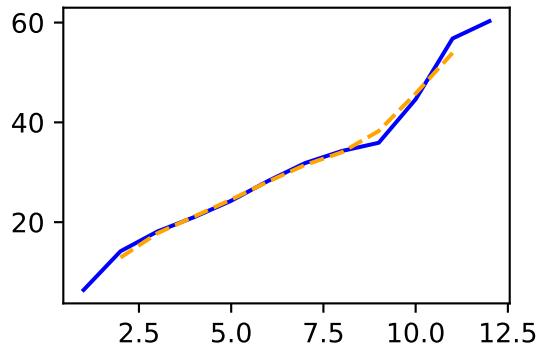
(b) Transient later joiner developers (n=119)



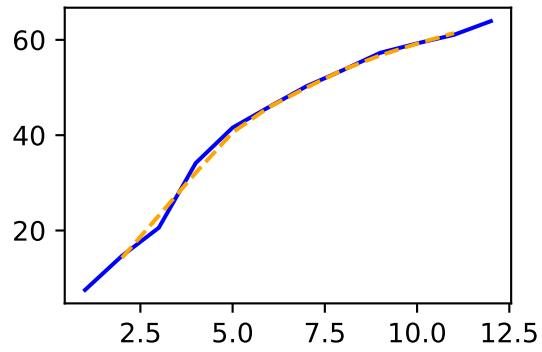
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 32: Repository: 14. A smooth moving of total packages touched on average each month of a year.

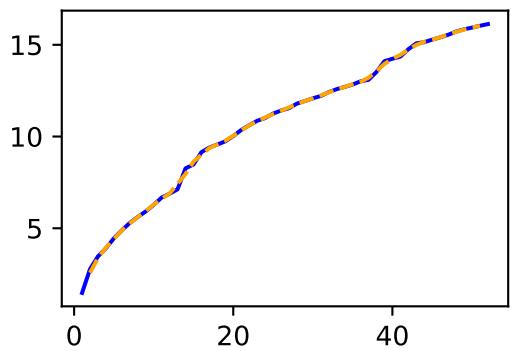
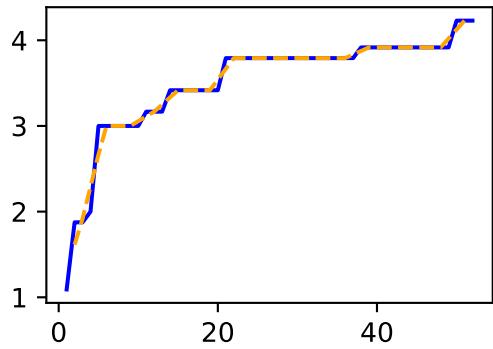
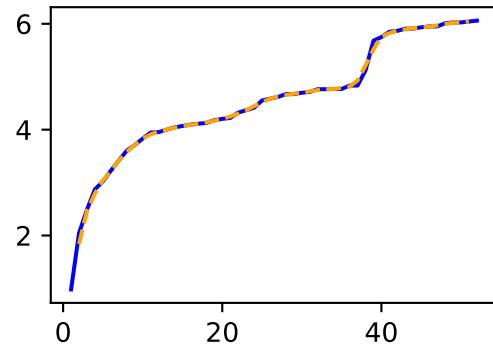


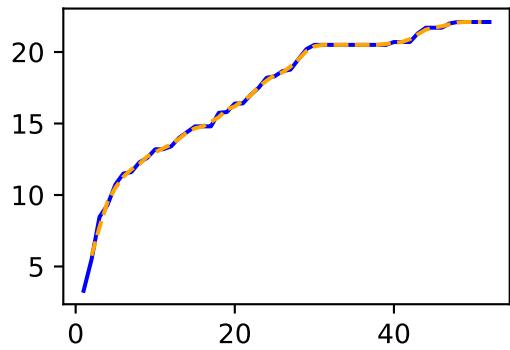
Figure 33: Repository: 14. A time series of the average (mean) total packages touched (y-axis) against the number of weeks (x-axis) for all developers ($n=154$).



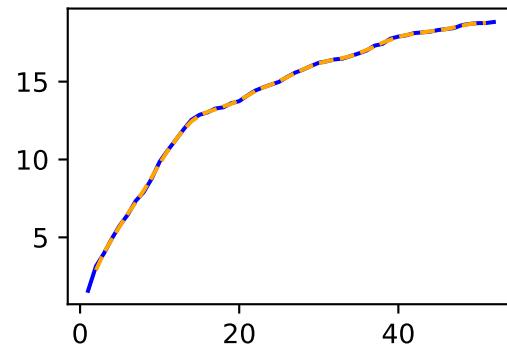
(a) Transient founder developers (n=8)



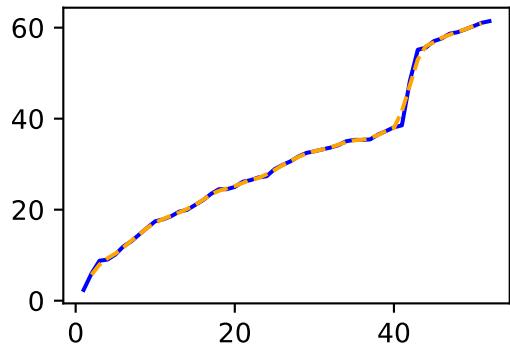
(b) Transient later joiner developers (n=119)



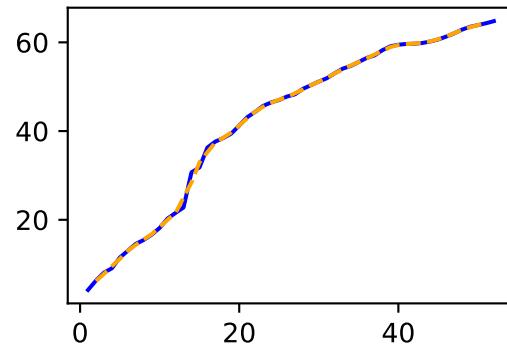
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 34: Repository: 14. A smooth moving of total packages touched on average each week of a year.

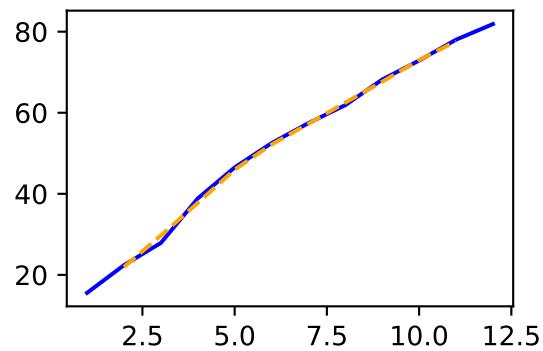
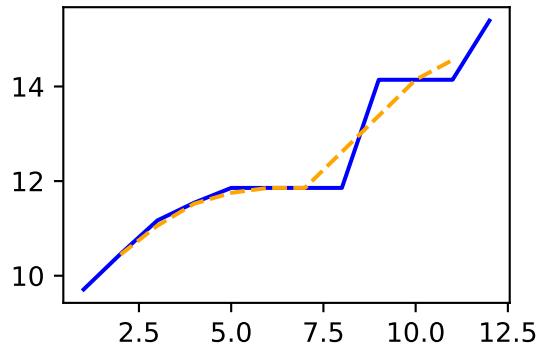
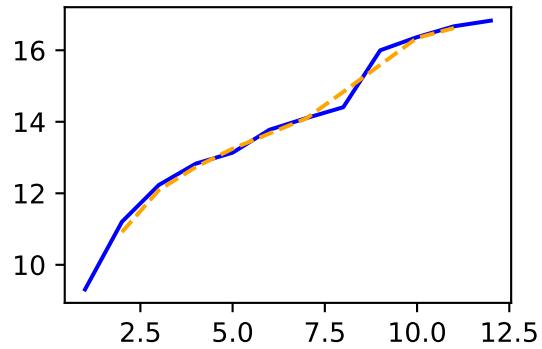


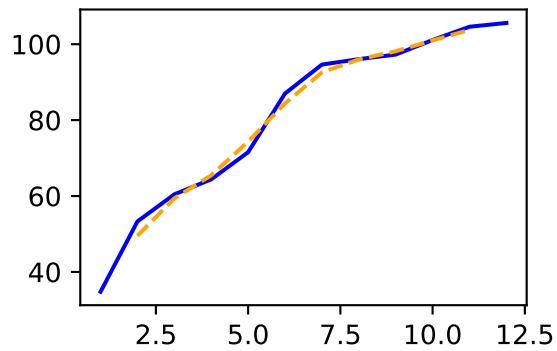
Figure 35: Repository: 14. A time series of the average (mean) total classes touched (y-axis) against the number of months (x-axis) for all developers (n=154).



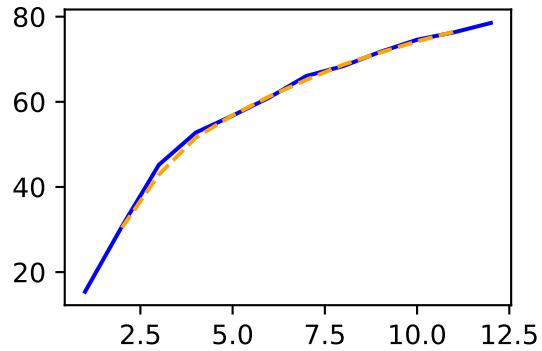
(a) Transient founder developers (n=8)



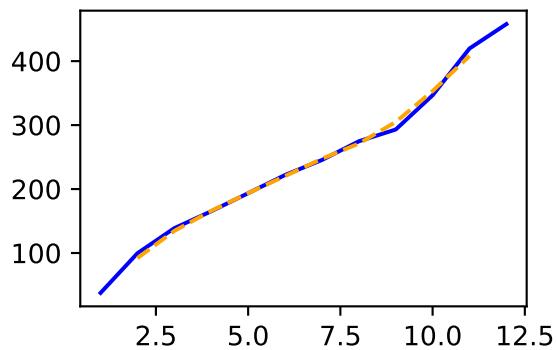
(b) Transient later joiner developers (n=119)



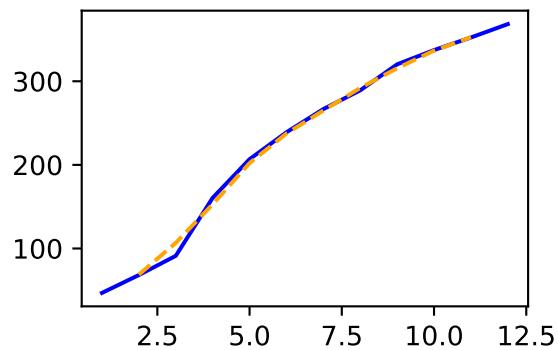
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 36: Repository: 14. A smooth moving of total classes touched on average each month of a year.

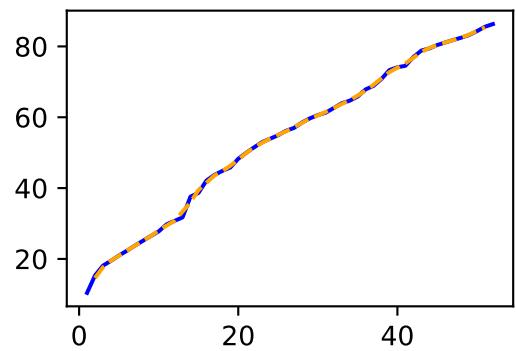
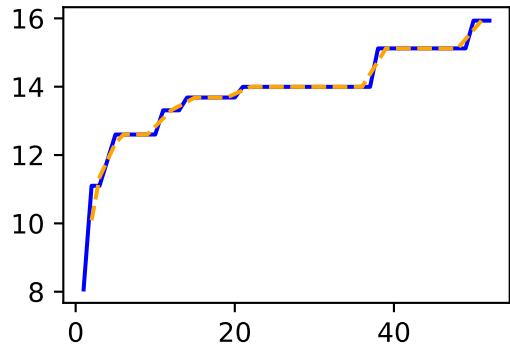
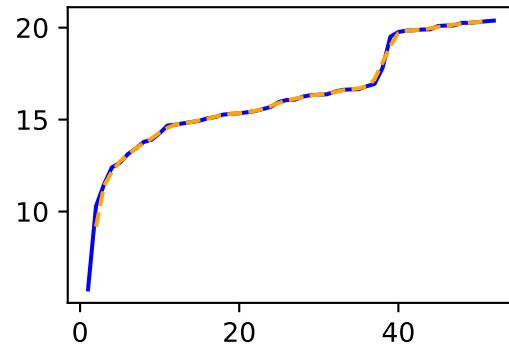


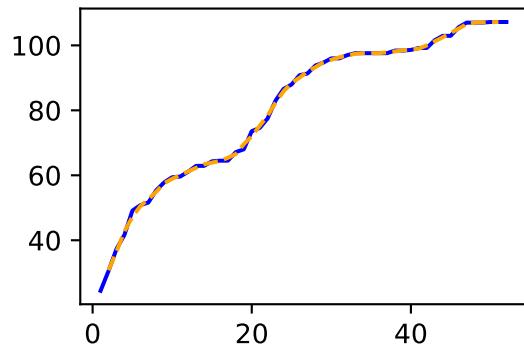
Figure 37: Repository: 14. A time series of the average (mean) total classes touched (y-axis) against the number of weeks (x-axis) for all developers ($n=154$).



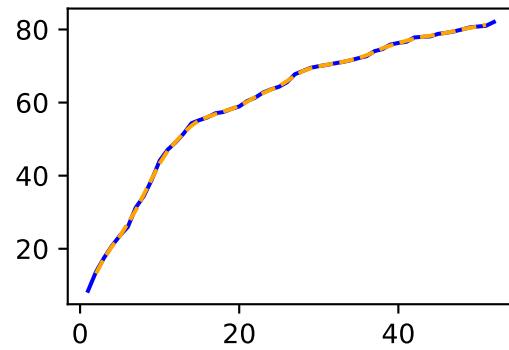
(a) Transient founder developers ($n=8$)



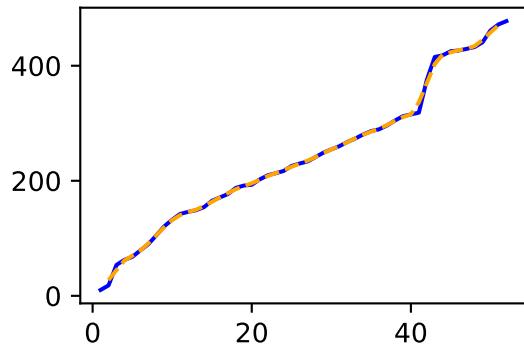
(b) Transient later joiner developers ($n=119$)



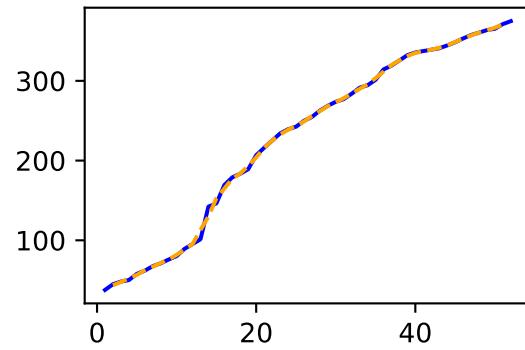
(c) Moderate founder developers ($n=5$)



(d) Moderate later joiner developers ($n=85$)



(e) Sustained founder developers ($n=6$)



(f) Sustained later joiner developers ($n=21$)

Figure 38: Repository: 14. A smooth moving of total classes touched on average each week of a year.

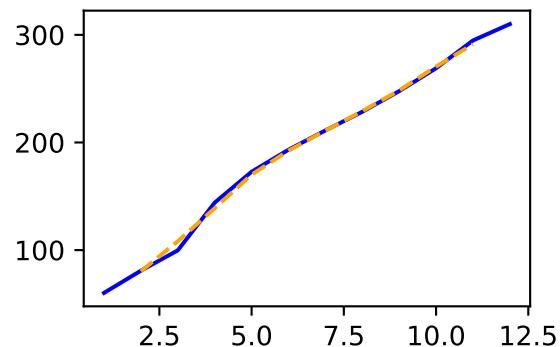


Figure 39: Repository: 14. A time series of the average (mean) total methods touched (y-axis) against the number of months (x-axis) for all developers (n=154).

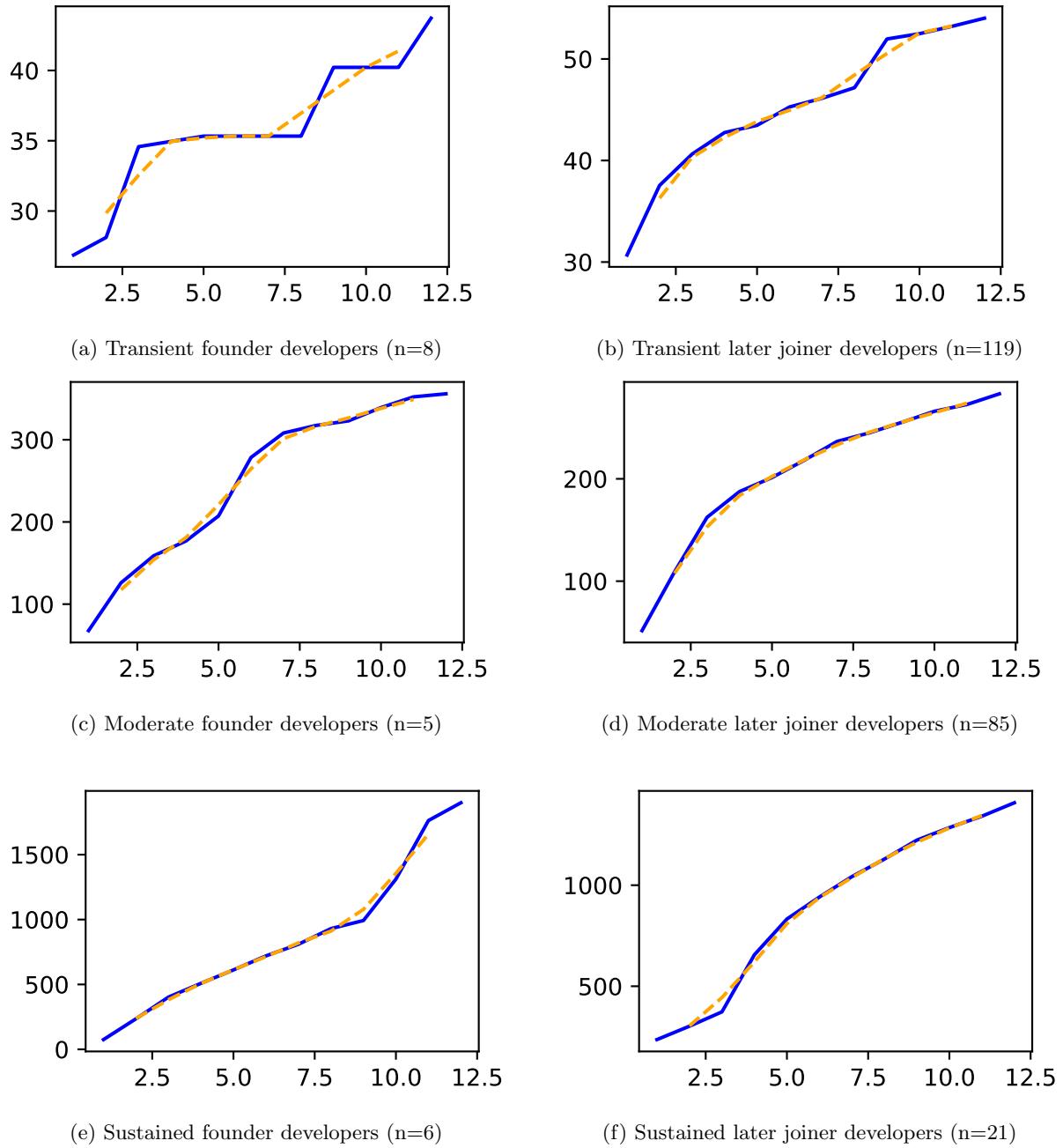


Figure 40: Repository: 14. A smooth moving of total methods touched on average each month of a year.

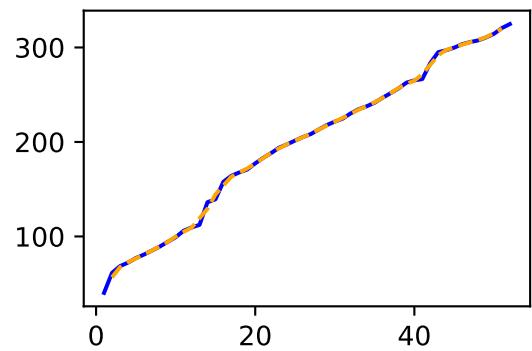
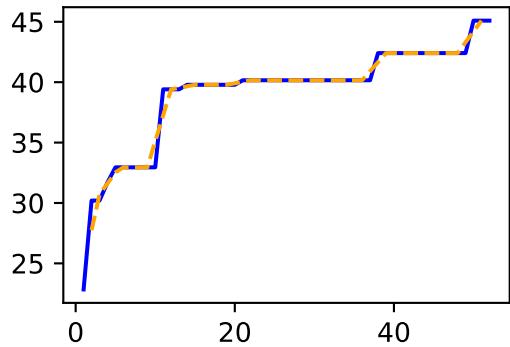
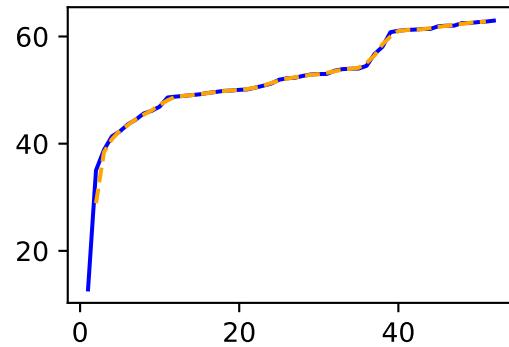


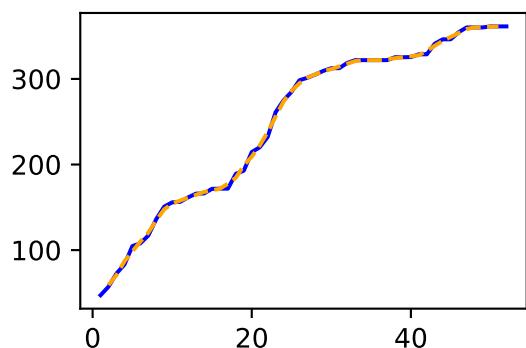
Figure 41: Repository: 14. A time series of the average (mean) total methods touched (y-axis) against the number of weeks (x-axis) for all developers (n=154).



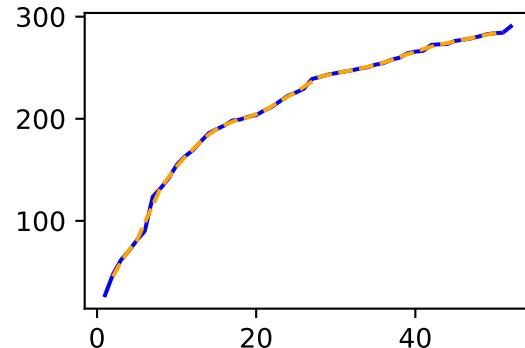
(a) Transient founder developers (n=8)



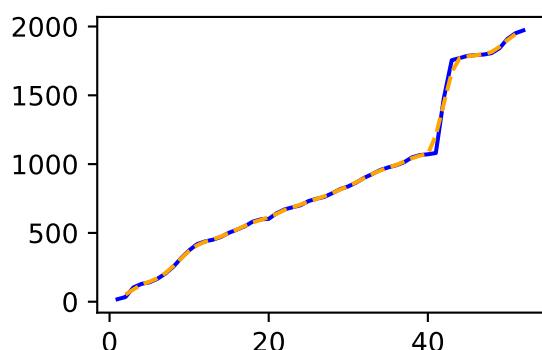
(b) Transient later joiner developers (n=119)



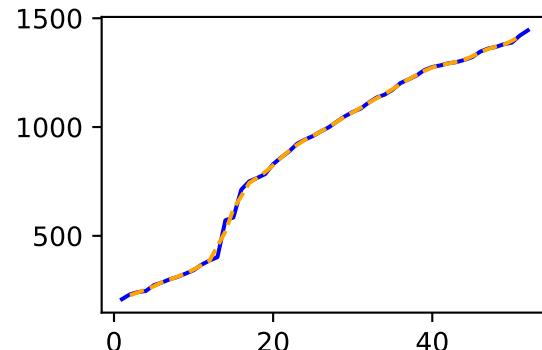
(c) Moderate founder developers (n=5)



(d) Moderate later joiner developers (n=85)



(e) Sustained founder developers (n=6)



(f) Sustained later joiner developers (n=21)

Figure 42: Repository: 14. A smooth moving of total methods touched on average each week of a year.

16.2 Smooth moving average - 21 For packages touched for each unit

A smooth moving average packages touched on average each period the number of commits to components touched.

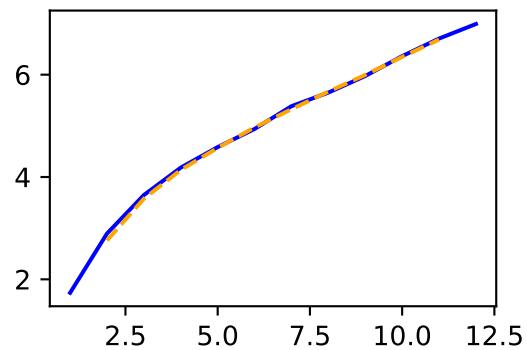


Figure 43: Repository: 21. A time series of the average (mean) total packages touched (y-axis) against the number of months (x-axis) for all developers (n=204).

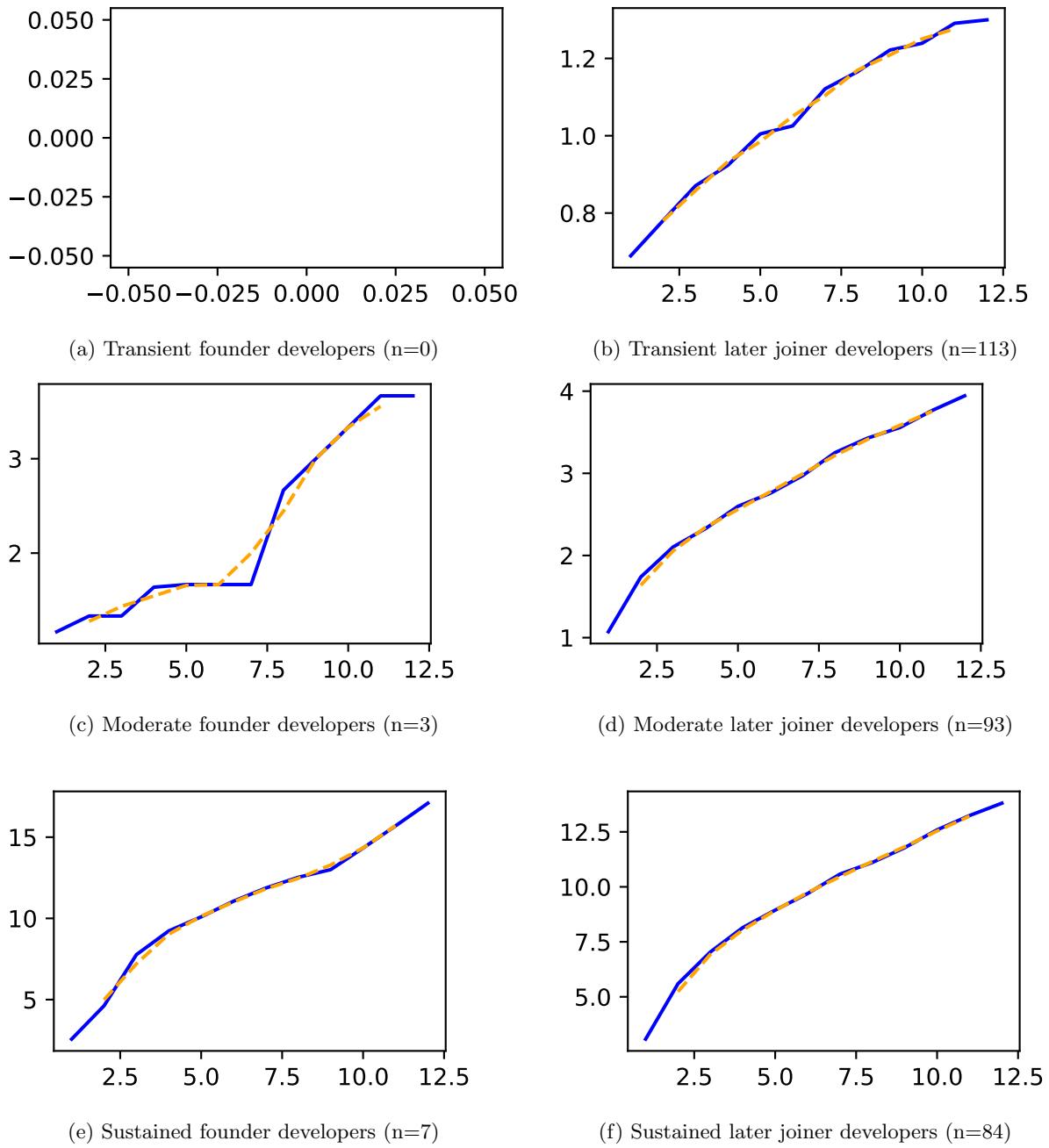


Figure 44: Repository: 21. A smooth moving of total packages touched on average each month of a year.

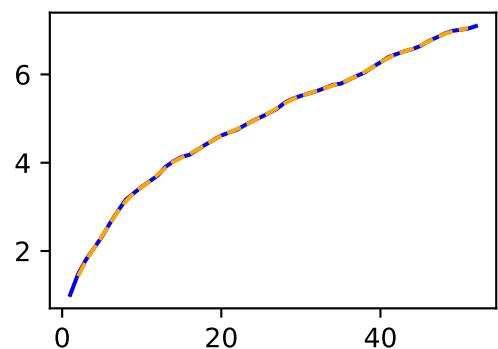


Figure 45: Repository: 21. A time series of the average (mean) total packages touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).

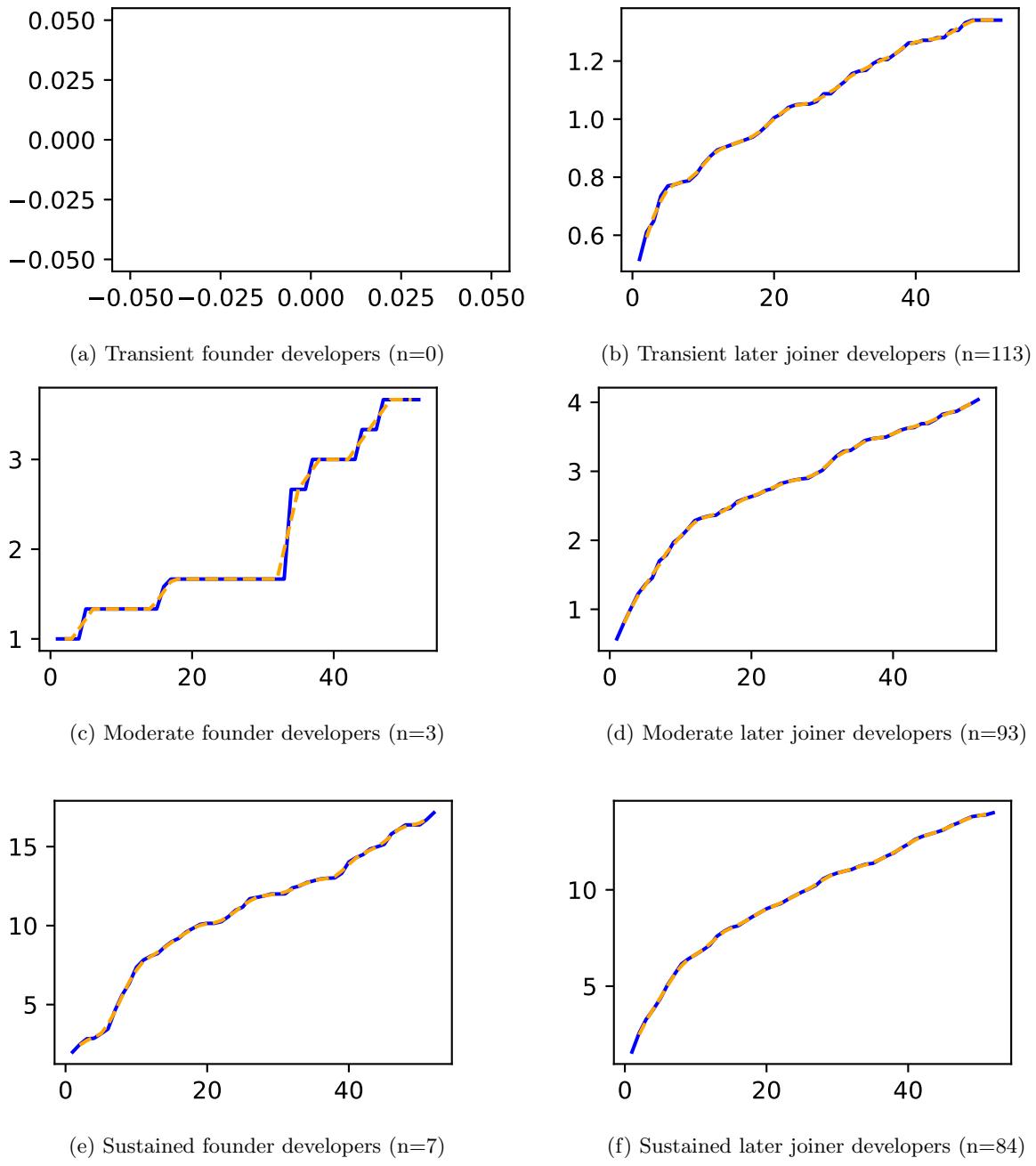


Figure 46: Repository: 21. A smooth moving of total packages touched on average each week of a year.

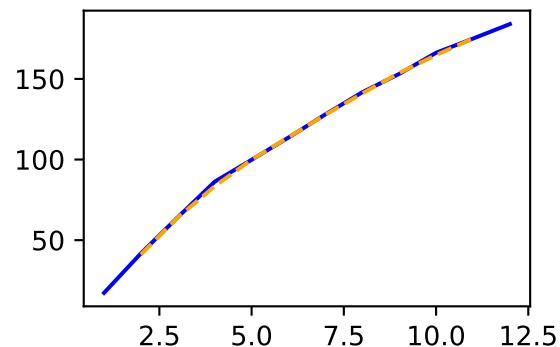


Figure 47: Repository: 21. A time series of the average (mean) total classes touched (y-axis) against the number of months (x-axis) for all developers (n=204).

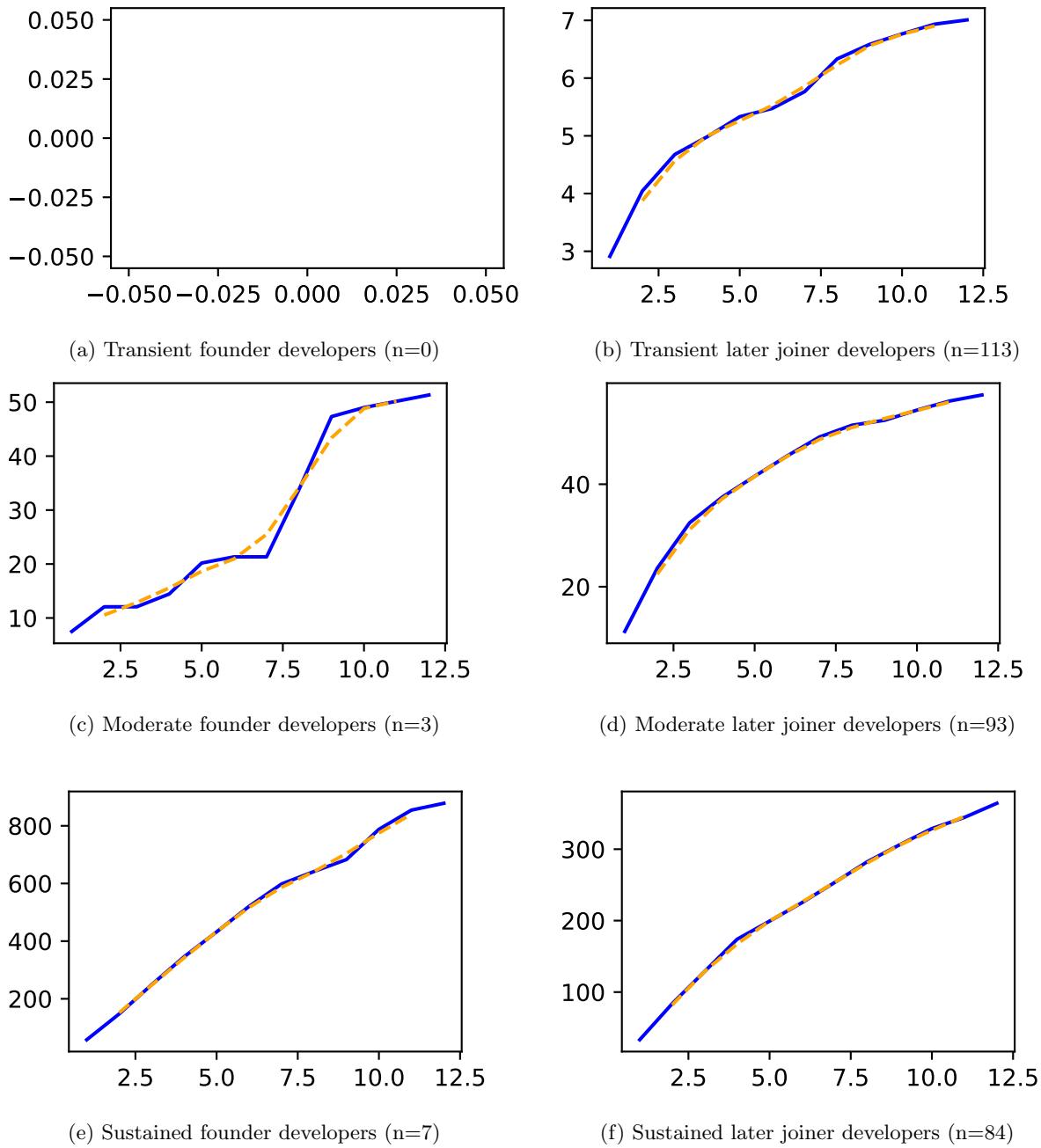


Figure 48: Repository: 21. A smooth moving of total classes touched on average each month of a year.

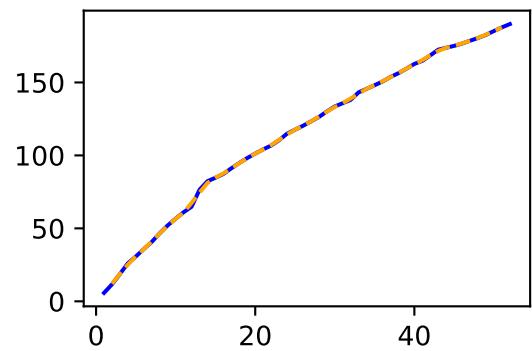


Figure 49: Repository: 21. A time series of the average (mean) total classes touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).

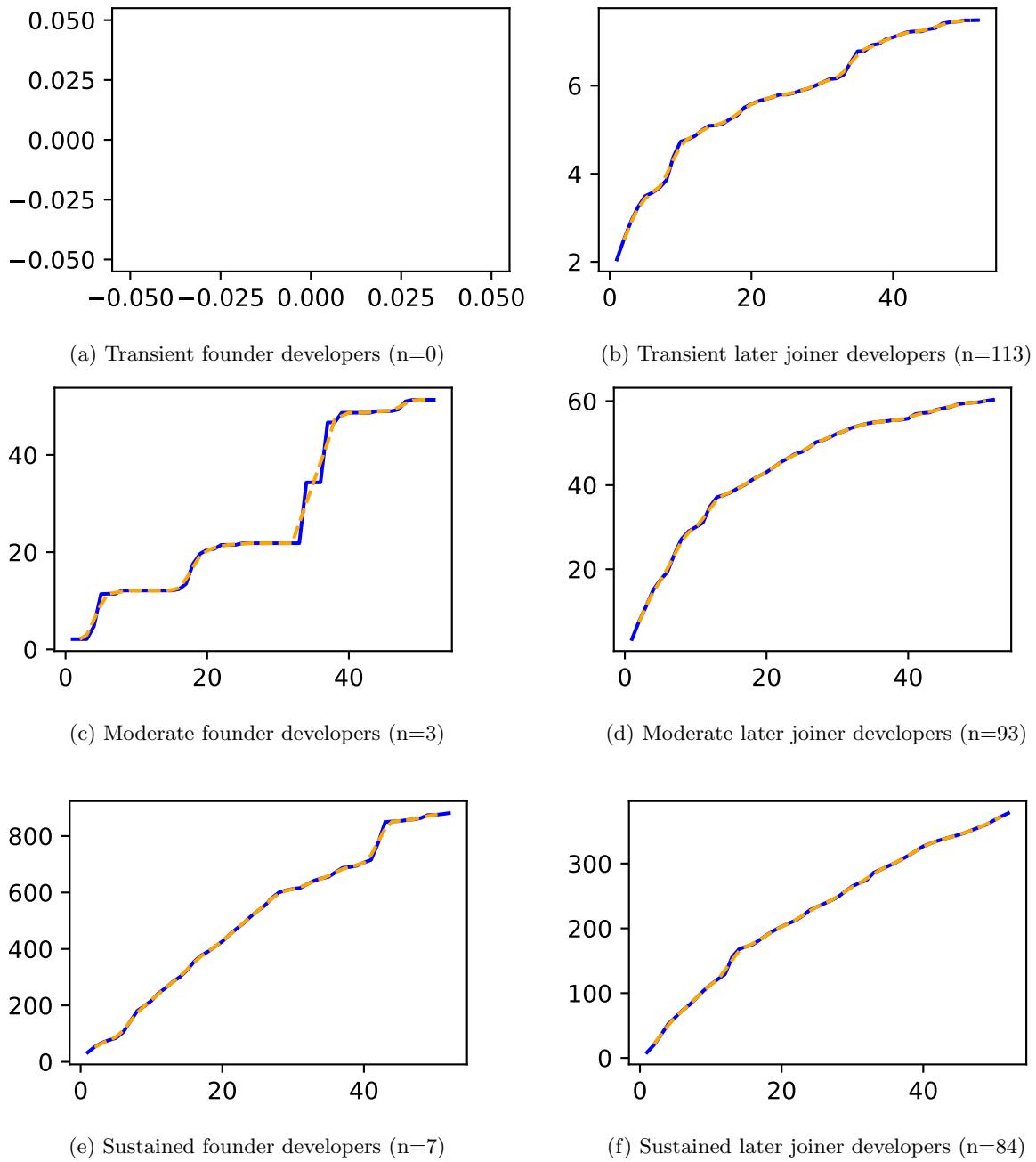


Figure 50: Repository: 21. A smooth moving of total classes touched on average each week of a year.

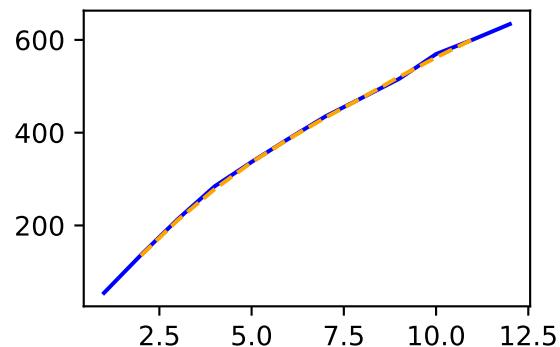


Figure 51: Repository: 21. A time series of the average (mean) total methods touched (y-axis) against the number of months (x-axis) for all developers (n=204).

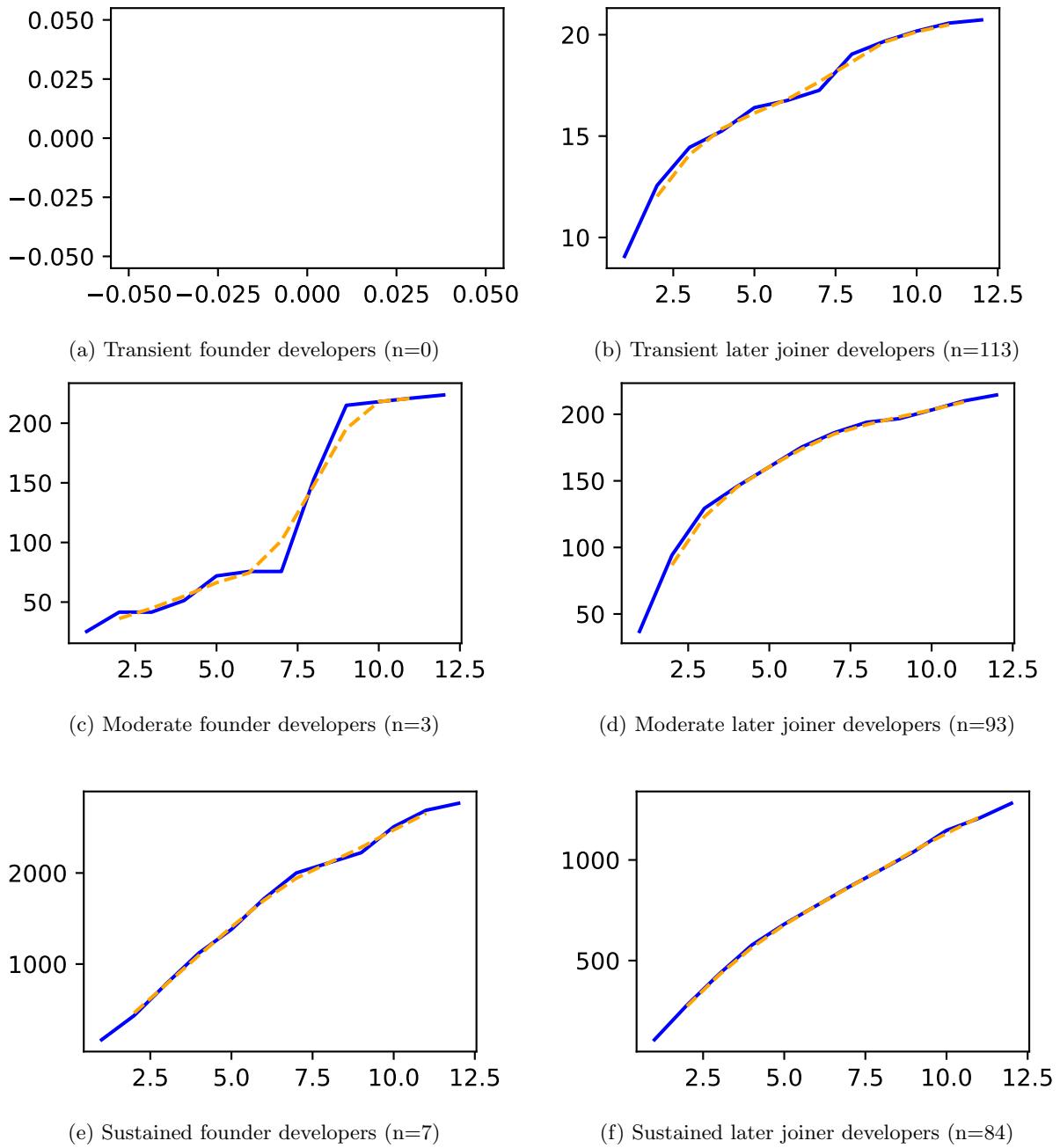


Figure 52: Repository: 21. A smooth moving of total methods touched on average each month of a year.

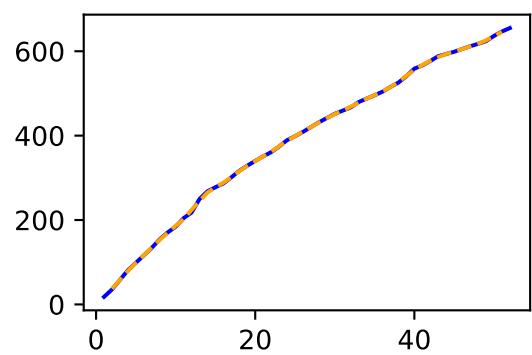


Figure 53: Repository: 21. A time series of the average (mean) total methods touched (y-axis) against the number of weeks (x-axis) for all developers (n=204).

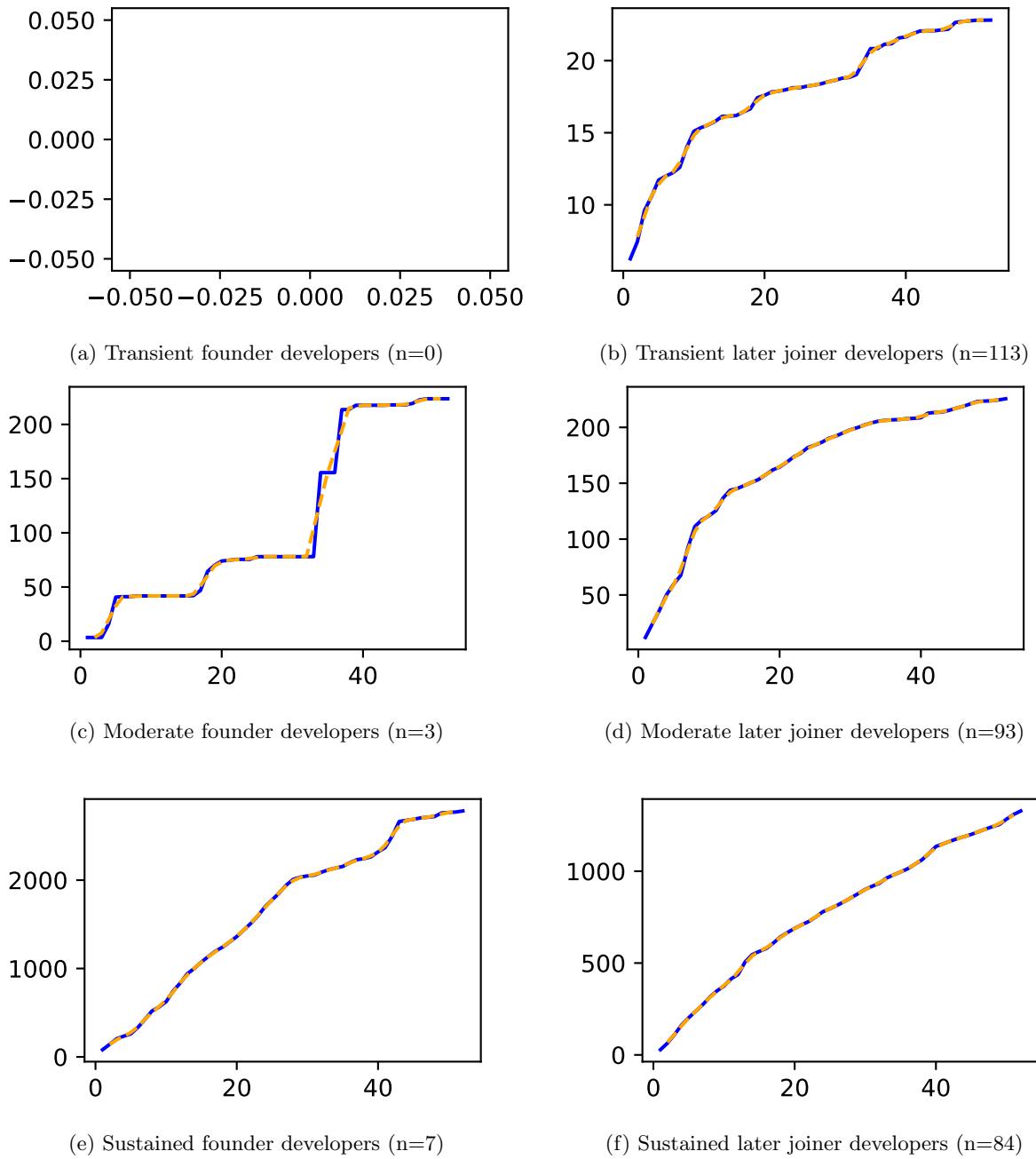


Figure 54: Repository: 21. A smooth moving of total methods touched on average each week of a year.

17 Repository pull request - Pull requests

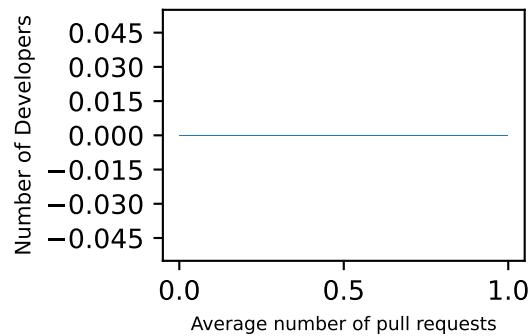


Figure 55: Average number of pull request for 0 developers.